Construct Validation of TOEFL-iBT (as a Conventional Test) and IELTS (as a Task-based Test) among Iranian EFL Test-takers' Performance on Speaking Modules

Elham Zahedkazemi

Department of English, Science and Research Branch, Ahvaz, Iran

Abstract—**The aim of the present research is to examine the construct validation of the speaking sections of two internationally known language proficiency examinations, namely IELTS and TOEFL iBT. IELTS (International English Language Testing System) and TOEFL iBT (Test of English as a Foreign Language, Internet-based Test) are two common examples of high-stake standardized tests which play a critical role in determining the future life of their candidates. This research seeks to scrutinize how IELTS and TOEFL iBT tap the same construct validity on the speaking proficiency of their candidates and moreover about the similarity of results for both IELTS and the TOEFL speaking tests. Participant in this study were 30 EFL students at high level of education in Khuzestan. Finding reveal that the IELTS and the TOEFL speaking tests share both differences and mostly similarities in terms of measuring a candidates' speaking ability. Different analyses yielded highly similar results: two tests are significantly correlated, classification of agreement between test-taker 'performance on both speaking test was very close. The study considers both the theoretical and practical implications of these findings for Iranian stakeholders in particular and for test users in general.**

Index Terms—**construct validation, speaking test, IELTS, TOEFL iBT**

I. INTRODUCTION

In today's competitive and technological world, students need to develop effective skills in mastering the English language in order to communicate ideas, concepts, scientific and technological findings successfully (grad.uprm.edu, 2010). One of the most important abilities to develop is the oral proficiency. Oral proficiency testing has become one of the most central topics in both language teaching and language testing; more particularly with the advent of communicative language teaching, the role of speaking ability has become more important (Shirinzadeh Aghdam and Farahani, 2012).

Conventional testing and Task-based assessment are two very different methods of assessing EFL learners at any level of speaking (Abrams and Madaus, 2003). Conventional testing refers to assessments that are traditional and old assessments (Moshman, 2011). Traditional assessments are the conventional methods of testing which usually produce a written document, such as a quiz, exam, or paper. Standardized tests, most state achievement tests, and high school graduation examinations are also examples of conversional or traditional assessments (ETS, 2012). While traditional assessment relies on indirect or proxy item-efficient, simplistic substitutes from which we think valid inference can be made about the student performance at those students challenges (Wiggins and Grant, 1990); Task-based language assessment (TBLA) grows from the observation that mastering the grammar and lexicon of a language is not sufficient for using a language to achieve ends in social situations (Mislevy, 2002).

A good example of conventional testing is TOFEL iBT. The Internet-based test (iBT) is the latest version of TOEFL, and it was launched in 2006 following more than a decade of validation activities intended to support the design and the proposed test score interpretations and uses of the test (Fulcher, 2003). The speaking construct assessed in the iBT was defined in terms of the knowledge and control of important phonological and syntactic features, vocabulary, and discourse patterns encountered in academic contexts (Cyril, 2007). On the other hand IELTS as a sample of task-based test in this study stand apart from other through its face-to-face speaking assessment. It is personal, interactive and as close to real-life situation as you can get. Institutions can rely on IELTS to assess candidates' ability to communicate in real-life situation, not just respond to pre-recorded prompts (Stoynoff, 2009). Among the four language skills included in these two exams (the IELTS and the TOEFL iBT), speaking, because of measurement difficulties, has received less attention in the available literature. This is the same in most language schools and universities in Iran. Plus measurement difficulties; Another problem is that even if speaking module has been studied in many respects, but its construct validity argument is nevertheless seriously compromised and compared in these two wide spread tests (Brown and Taylor, 2006). Construct validity is one of the most central concepts in language testing field. Researchers generally establish the construct validity of a measure by correlating it with a number of other measures and arguing from the

pattern of correlations that the measure is associated with these variables in theoretically predictable ways (Cronbach and Meehl, 1955, p. 288).

Based on language testing research center (ITRC) statistics in 2014 the IELTS and the TOEFL iBT tests have been used widely and in the world as EFL language proficiency tests for admission for tertiary education in English speaking countries. Generally, scores of the two tests can be used interchangeably as the number of institutions in the world accepting both scores (of the IELTS and the TOEFL iBT) as evidence about as English language proficiency has increased greatly. The issue is whether we can make the same inference about test-takers' language proficiency based on the scores the obtained on each test. It is therefore of great importance to investigate and compare the validity and particularly the construct validity of two tests in general and of the speaking sub-test in particular.

Research Questions

This study will pose the following research questions:

1. Do the IELTS and the TOEFL iBT speaking test measuring the same speaking ability of the test-takers?

2. Could we draw the same inferences about candidates' abilities based on each test they take?

Hypotheses

H01. The IELTS and the TOEFL iBT speaking test measuring the same speaking ability of the test-takers.

H02. We could draw the same inferences about candidates' abilities based on each test they take.

II. METHODOLOGY

Participants

The participants of this study consisted of 50 Iranian EFL learners including MA students of both genders majoring in English teaching, in Islamic Azad University of Khuzestan, Science and Research branch; and also EFL learners at high level of proficiency in different language institutions in Dezful who all sat for a homogeneity test. Their ages ranged between 20 to 34. Out this population, 30 test-takers were selected based on the results of their performance on a given homogeneity test which was the Quick Placement Test of Oxford University Press and University of Cambridge Local Examinations Syndicate (2001). Then an oral interview was conducted to test their speaking ability, by two raters .The learners based on their proficiency level were divided to two groups of proficiency: intermediate and upper-intermediate.

Material

The material employed in this study consisted of a language proficiency test (OPT), an Oral interview test, IELTS test held in Australia in January 2014 and TOEFL iBT Longman speaking test version 2013. These materials were chosen for several reasons. Firstly, it was impossible to obtain genuine tests due to test security of both the IELTS test and the TOEFL iBT test. Secondly the IELTS test administered to the test-takers was held on January, 2014 that was the closet and newest one to the real IELTS test compared to any other commercial IELTS practice test available. Similarly, TOEFL iBT Longman software published by Pearson Long- man and collected by Deborah Philips (2013) is one of the most authentic tests available and is the closest test to the real TOEFL iBT in comparison with other commercial iBT tests. Researcher has recorded the participants' voice during the speaking exams of both tests. The TOEFL exam program had the option to save the candidates' voice and the IELTS speaking file saved separately by the researcher. Later on, test-takers speaking recordings were saved in a particular folder for evaluations and scoring.

Instrumentation

A language proficiency test (OPT) to homogenize the prospective test-takers for the study. Then due to this reason why OPT exam lacks a speaking section; an Oral interview test was administered to. Researcher implemented the same oral interview which Shirinzade Aghdam and Farahai (2012) used in their study on speaking skill developed by Underhill (1987) and Brown (2004). The main instruments utilized was consisted of the IELTS test held in Australia in January 2014, and TOEFL iBT Longman software version 2013, which is a practice test for those getting prepared for real TOEFL iBT exam.

Procedure

As it was mentioned earlier (part 3.1), from among a population of 50 people 30 participants were selected to enter the present study based on their performance on the proficiency test– the Intermediate and Upper-intermediate levels – and each level of test-takers sat once for the IELTS speaking test and once for the TOEFL iBT speaking test (see sections 3.1 and 3.2 in this chapter). Participants were first provided with the required information about IELTS and TOEFL iBT's distinguishing features and characteristics on speaking part during the two sessions. The IELTS speaking test was administered to test-takers by a non-practicing IELTS examiner; the speaking interviews of each test-taker were recorded for later analyses and scoring. A non-practicing certificate IELTS examiner scored speaking modules of both tests for each level of test-takers proficiency. Two main functions were performed by the certificate IELTS examiner: rating the test-takers voice samples, and providing the researcher with information about speaking interview procedures of the participants. Once again as it was already mentioned in the instrument section TOFEL iBT speaking, the sub-test had two types of tasks: independent and integrated Each of these tasks was rated from 0 to 4. Regarding IELTS, the speaking scale is composed of ten levels, ranging from 0 to 9. Each of test-takers' performance on his/her speaking test was scored on specified criteria for each test. Regarding TOEFL iBT, the speaking exam, raters or interviewers did assess test-takers' speech samples from four the following criteria: General description, Delivery,

Language use and Topic development; and for the IELTS speaking test, test-takers performances were assessed in terms of the following four criteria: Vocabulary, Grammar, Fluency and pronunciation. By the end of the two tests, for each of the 30 participants, researcher could collect 8 sets of scores on each band of their speaking tests. 8 correlation analyses were carried out to examine the relationship between the different parts within and between this two speaking tests. The correlation between the IELTS and the TOEFL iBT speaking tests was calculated to see if the same conclusion could be reached out about candidates' ability on each test (using Pearson correlation from the SPSS program version 18, see chapter four for the results).

III. RESULT AND DISCUSSION

Research question number 1: The IELTS and the TOEFL iBT speaking tests maeasure the same speaking ability of the test-takers?

- Descriptive statistics of the raw scores

The basic descriptive statistics of the IELTS and TOEFL iBT speaking test-scores of 30 test-takers both in Intermediate and Upper-intermediate groups are presented in Table 1 and Table 2 below.

TABLE 1:
DESCRIPTIVE STATISTICS OF THE IELTS AND THE TOEFL iBT SPEAKING SCORES IN UPPER-INTERMEDIATE GROUP

Upper-intermediate	The IELTS speaking test (N=12)	The TOEFL iBT speaking test (N=12)
Mean	23.73	23.65
Median	23.41	23.10
Std.Deviation	2.55	3.29
Minimum	20	19.5
Maximum	28	29

After converting all test-takers' scores to the same scale – in this research, researcher took TOEFL iBT scale scores as criterion based on tables of conversion documented by ETS organization. It can be seen that for Upper-intermediates, mean and median values of the IELTS speaking scores were highly similar at 23.73 and 23.41 respectively). Similar to the IELTS speaking scores, the mean and the median values of the TOEFL iBT speaking test were quite close together at 23.65 and 23.10, respectively, indicating a nearly normal distribution.

TABLE 2:
DESCRIPTIVE STATISTICS OF THE IELTS AND THE TOEFL iBT SPEAKING SCORES IN INTERMEDIATE GROUP

Intermediate	The IELTS speaking test (N=18)	The TOEFL iBT speaking test (N=18)
Mean	19.39	18.78
Median	19.00	18.38
Std.Deviation	2.21	2.97
Minimum	17	16.5
Maximum	21	22

Regarding to Table 2 the similar results as for the Upper-intermediate group were driven for the Intermediate group of test-takers (mean: 19.39 and median: 19) showing a normal distribution of test takers. Similar to the IELTS speaking scores, the mean and the median values of the TOEFL iBT speaking test were also nearly close together at 18.78 and 18.38, respectively that show a normal distribution among test-takers.

-Pearson Correlation coefficient of test-takers' scores

This is also reflected in the Pearson Correlation coefficient table 10 that in Intermediate group of test-takers Correlation coefficient is 0.701 (r =.701, p<.001) which signifies that generally those who did well in the IELTS speaking test tended to do well in the TOEFL iBT speaking test and vice versa. The result for Upper-intermediate group (r = .870, p<.001) again proves that test-takers who did well on the IELTS speaking test are likely to do well on TOEFL iBT speaking test and vice versa. However, r = .701 indicates only a moderate correlation between two speaking tests in Intermediate group, while an r = .870 for Upper-intermediate group (Table 1) of test-takers indicated that there is stronger correlation between two the speaking tests in this group. These findings are a source of evidence in support of Hypothesis 1: The IELTS and the TOEFL iBT speaking test measure the same speaking ability of the test-takers.

Research question number 2: Could we draw the same inferences about candidates' abilities based on each test they take?

- Analyzing the relationship between test-takers' achievements

In order to analyze the relationship between test-takers' achievements – as measured by their scores on the two speaking tests of the IELTS and TOEFL iBT - and their abilities in speaking skill, their performance on each speaking exam in greatest details are presented in Tables 3 and 4 for both groups of Intermediate and Upper-intermediate.

TABLE 3:
COMPARISON OF IELTS AND iBT SPEAKING TESTS IN TERMS OF THEIR CRITERIA (UPPER- INTERMEDIATE GROUP)

Upper-Intermediate group (IELTS) N=12				Upper-Intermediate group (TOEFL iBT) N=12			
Speaking Criteria	Mean	Std.Dev	Correlation Coefficient	Speaking Criteria	Mean	Std.Dev	Correlation coefficient
Vocabulary	24.08	3.20	0.95	General Description	23.90	3.74	0.94
Grammar	23.92	3.37	0.99	Delivery	23.98	2.65	0.86
Pronunciation	24.25	2.96	0.91	Language use	24.65	3.78	0.93
Fluency	22.67	4.16	0.98	Topic Development	22.09	3.69	0.89

TABLE 4:
COMPARISON OF IELTS AND iBT SPEAKING TESTS IN TERMS OF THEIR CRITERIA (INTERMEDIATE GROUP)

Intermediate group (IELTS) N=18				Intermediate group (TOEFL iBT) N=18			
Speaking Criteria	Mean	Std.Dev	Correlation Coefficient	Speaking Criteria	Mean	Std.Dev	Correlation coefficient
Vocabulary	19.39	2.06	0.89	General Description	18.55	1.92	0.85
Grammar	20.39	1.82	0.82	Delivery	19.50	2.12	0.78
Pronunciation	19.06	2.07	0.88	Language use	21	2.08	0.91
Fluency	18.94	1.39	0.76	Topic Development	16.12	1.78	0.83

As it can be seen in Table 3 , for Upper-intermediate group, test-takers' performance in the IELTS group were closer in all four areas of vocabulary, grammar, pronunciation and fluency, though the IELTS 'test-takers best performance was in the area of grammar (mean= 23.92, r = 0.99), test-takers mostly could use a range of complex structure with required flexibility with error-free sentences and few grammatical mistakes, while their performance in area of fluency was found to be weaker than the other areas of their speaking skill (Mean=22.67, r = 0,98). As for the TOEFL iBT test takers, raw scores revealed that while their performance as a whole was quiet similar to their performance for the IELTS speaking test (IELTS: 23.7, iBT: 23.65), the range and mean of their scores in four areas were not close.

Investigation for Intermediate group of test-takers in Table 4 revealed that test-takers' highest and lowest performances were in grammar and fluency areas (mean= 20.39, r = 0.82 and mean= 18.94, r = 0.76, respectively) which are truly similar to Upper-intermediate test-takers' performance on the highest and weakest areas of the IELTS speaking test. The correlation between test-takers 'performances in the four areas of vocabulary, grammar, pronunciation and fluency for both groups and the IELTS speaking test as a whole were found to be strong (r = 0.95 for Upper-intermediate and r = 0.83 for Intermediate group). This correlation appeared to be stronger for Upper-intermediate group than the Intermediate group. Likewise on performance for the TOEFL iBT Intermediate group, the result showed that test-takers performed strong in "language use" area and really weak in "topic development" area (mean= 21, r = 0.91 and mean= 16.12, r = 0.83, respectively). These results appeared to be consistent with the TOEFL iBT 'Upper-intermediate group performance. The correlation between test-takers 'performances in the four areas of general description, delivery, language use and topic development for both TOEFL iBT groups and speaking test as a whole were found to be strong (r = 0.89 for Upper-intermediate and r = 0.79 for Intermediate group). And finally a summary of basic information in the IELTS and the TOEFL iBT speaking test performance raw scores across the two test-takers groups is shown in Table 5 below in order to come up with some basic results and conclusions related to the first hypothesis in this study.

TABLE 5:
DESCRIPTIVE STATISTICS OF THE IELTS AND THE TOEFL iBT SPEAKING RAW SCORES FOR EACH TEST IN TOTAL COMPARISON.

Group	Mean	Median	Std.Dev	Correlation	Minimum	Maximum
IELTS (Upper-intermediate)	23.73	23.41	2.91	0.95	20	28
TOEFL iBT (Upper-intermediate)	23.65	23.10	3.29	0.89	19.5	29
IELTS (Intermediate)	19.39	19.04	2.21	0.83	17	21
TOEFL iBT (Intermediate)	18.78	18.38	2.97	0.79	16.5	22
IELTS and iBT (Upper-intermediate)	23.69	23.50	3.25	0.87	19.5	29
IELTS and iBT (Intermediate)	19.08	18.88	3.72	0.70	16.5	22
IELTS	21.56	21.24	3.21	0.89	17	28
TOEFL iBT	20.67	20.39	4.01	0.80	16.5	29

Total comparison for the IELTS and the TOEFL iBT' scores of the test-takers

In total comparison for IELTS and the TOEFL iBT, as it is demonstrated in the raw scores of both groups of test-takers, firstly, the maximum score of the IELTS test-takers was very close to the TOEFL test-takers (17 and 16.5, respectively), likewise the minimum scores of both groups were at the close level (28 and 29, respectively). In addition, the median of the IELTS speaking test and the TOEFL iBT was at the minimum level of difference (21.24 and 20.39). Thus the primarily finding is that the IELTS test-takers and the TOEFL iBT test-takers both performed at the very close level of proficiency in their speaking tests. Evidence from the Table 5 showed that the mean score of the IELTS speaking test was higher than the mean score of the TOEFL iBT speaking test (21.56 vs. 21.21). In contrast, standard deviation of the IELTS speaking test was smaller than the TOEFL iBT speaking test (3.21 vs. 4.01) which means the

IELTS test-takers seemed to perform not only better on the test they were attended to but also in more homogeneous way than the TOEFL iBT test-takers. To examine if these differences between the two groups and across the two tests as whole were statistically significant, an independent t-test was used after conforming that the scores distributions for each group were normally distributed, meeting the requirements of an independent t-test (see Hatch and Lazaraton, 1991).

TABLE 6:
THE RESULTS OF THE INDEPENDENT SAMPLE T-TEST FOR COMPARISON OF
THE TWO GROUPS OF TEST-TAKERS ON IELTS AND TOEFL iBT (INTERMEDIATE AND UPPER-INTERMEDIATE)

IELTS and TOEFL Upper-intermediate N=12	Levene's test for equality of Variances		t-test for Equality of Mean					95% confidence interval of Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean differences	Std. Error differences	Lower	Upper
Equal variance assumed	.241	.628	.525	22	.605	.666	1.268	-1.964	3.297
Equal variance not assumed			.525	21.691	.605	.666	1.268	-1.966	3.300
IELTS and TOEFL iBT intermediate N= 18									
Equal variance assumed	32.39	.000	-1.270	34	.048	-1.286	1.0123	-3.343	.771
Equal variance not assumed			-1.270	22.57	.017	-1.286	1.0123	-3.382	.810

The independent t-test showed a significant difference between the two speaking tests in Intermediate group with regard to their means of the IELTS and the TOEFL iBT speaking tests (19.39 vs. 18.78, t = -1.270, p <.05). In other words, the IELTS Intermediate group performed significantly better than the TOEFL Intermediate group. As for Upper-intermediate group, we could see there was no significant difference between the IELTS and the TOEFL speaking tests with the regard to their means logit scores on two tests (23.73 vs. 23.65, t = .525, p >.001).

Once again to examine if the difference between scores for two groups of the IELTS and the TOEFL iBT test-takers as a whole, was statistically significant, an independent t-test was run. The distributions of the IELTS and the TOEFL iBT speaking ability estimates of each group were first checked and a normal and close to normal distribution could be seen in the ability estimates of both groups meeting the requirements of an independent t-test.

TABLE 7:
THE RESULTS OF THE INDEPENDENT SAMPLE T-TEST FOR THE TWO GROUPS OF TEST-TAKERS
IN BOTH IELTS AND TOEFL iBT SPEAKING TESTS AS A WHOLE.

IELTS and TOEFL Upper-intermediate N=30	Levene's test for equality of Variances		t-test for Equality of Mean					95% confidence interval of Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean differences	Std. Error differences	Lower	Upper
Equal variance assumed	8.537	.005	-.484	58	.630	-.505	1.043	-2.592	1.582
Equal variance not assumed			-.484	51.123	.630	-.505	1.043	-2.598	1.588

The result for the independent t-test presented in Table 7 demonstrated that the difference between the IELTS and the TOEFL iBT speaking tests as a whole was not statistically significant in both their performance of the test-takers on tests and their ability towards the two speaking tests. There was also no significant difference between the IELTS and the TOEFL iBT speaking tests with regard to their means (21.56 vs. 21.21, t = -.484, p> .001). These results could suffice a piece of evidence to support the second sub-hypothesis in this study that: We could draw the same inferences about candidates' abilities based on each test they take.

IV. CONCLUSION AND IMPLICATIONS

In brief, all the analyses revealed that the IELTS and the TOEFL speaking tests share both differences and mostly similarities in terms of measuring a candidates' speaking ability. Different analyses yielded highly similar results: two tests are significantly correlated, classification of agreement between test-taker 'performance on both speaking test was very close, indeed for Upper-intermediate that was in a higher correlation, but in total comparison both the IELTS and the TOEFL iBT speaking tests could yield closely similar results for test-takers, and that could be a strong support for the research hypotheses. The current study is the first attempt made by the researcher to investigate and compare the construct validity of the IELTS and the TOEFL iBT speaking tests of its type in Iranian context. Finding of this investigation can provide useful information regarding Iranian test-takers' reactions to these two tests: they have highlighted that both functions as well as measures of their true speaking ability but have emphasized the value of test preparation.

- Implications of the study

1. The approach used in this study to compare the construct validity of two test is driven from two sources: what Messick (1989) and Kane (1992-2002) suggested for (a) data sources and analyses relevant to construct validation purposes, and (b), this researcher' s critique of what other researchers have done when comparing two tests. This multi-faceted approach has contributed richer and more varied information about the similarities and differences between two tests constructs than has been the case with pervious test comparison research.

2. Even if two test are widely used interchangeably for the same purpose of (in this case the admission of non-native English speaking students to tertiary education) and they are significantly, they cannot be assumed to be measuring the same construct at all levels as it was shown in first part of chapter three. Researcher' analysis of the test-takers' scores imply the developers of the IELTS should consider the value of creating a more academically oriented speaking test for the Academic Module in its battery.

3. Test preparation training is likely to have some effects on test performance – in this study there was a very short test preparation training about 3 sessions – especially when the test format is complex as has been demonstrated in the IELTS and the TOEFL iBT speaking tests for both groups of test-takers at each level. Therefore test-takers should be encouraged to take part in test preparation courses or at least make them familiar with the test format through practice before actually sitting for the test.

<div align="center">REFERENCES</div>

[1]   Abrams, L. M. & Madaus, G. F. (2003). The lessons of high-stakes testing. *Educational Leadership, 61*(3), 31-35.
[2]   Alderson, C. Clapham, C. & Wall, D. (1995). Language test construction and evaluation. New York: Cambridge University Press.
[3]   Aryadoust, V. (2011). Validity Arguments of the Speaking and Listening Modules of International English Language Testing System: A Synthesis of Existing Research. *Asian ESP Journal* Volume 7 - 2 Spring 2011, 25-34.
[4]   Bachman, L. F. (2006, April). Linking validity and use in educational assessments. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
[5]   Baleghizadeh, S. (2007). Speech acts in English language teaching. *Iranian Journal of Language Studies*, 1, 143–154
[6]   Berthold, M. (2011). Reliability of Quick Placement Tests: How much faith can we place on quick paper or internet based placement tests? *Australian Journal of Teacher Education*, 35(6), 1-23.
[7]   Birjandi, P. & Ahangari, S. (2008). Effects of task repetition on the fluency, complexity and accuracy of Iranian EFL learners' oral discourse. Retrieved December, 17, 2009, from: http://www.asian-efl-journal.com/site_map_2008.php.
[8]   Brown, H.D. (2006). Language assessment: Principles and classroom practices. New York: Pearson Education.
[9]   Brown, A. & Taylor, L. (2006). A worldwide survey of examiners' views and experience of the revised IELTS speaking test. *Research Notes*, *26*, 14-18.
[10]   Bygate, M. (2002). Speaking. In C., Roland and D., Nunan (Eds). *Teaching English to speakers of other languages.* Cambridge: Cambridge University Press.
[11]   Cohen, A. (1984). On taking language tests: What the students report. *Language Testing*, 1, 70-81.
[12]   Costa, A. L., & Kallick, B. (1992). Reassessing assessment. In A. L. Costa, J. A. Bellanca, & R. Fogarty, (Eds.), *If minds matter: A forward to the future, Volume II* (pp. 275-280). Palatine, IL: IRI/Skylight Publishing.
[13]   Criper, C. & A. Davies. (1988). Research Report 1 (i). ELTS Validation Project Report Cambridge: British Council/UCLES.
[14]   Cyril, J. W. & Saville, N. (2007). Using observation checklists to validate speaking-test tasks. *University of Cambridge Local Examinations Syndicate. Language Testing* 2002 19 (1) 33–56.
[15]   Ducasse, *A. & Brown, A. (2009).* The Role of Interactive Communication in IELTS Speaking and Its Relationship to Candidates' Preparedness for Study or Training Contexts. Www.Ielts.Org. May 17, 2011, 56-66.
[16]   Educational Testing Service (2007). TOEFL iBT reliability and generalizability of scores. Princeton, NJ: Educational Testing Service.
[17]   Fulcher, G. (2003). Testing second language speaking. London: Longman.
[18]   Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). TOEFL 2000 framework: A working paper (TOEFL Monograph Series Rep. No. 16). Princeton, NJ: ETS**.**
[19]   kim, H. (2002). Task-based Performance Assessment for Teachers: Key Issues to Consider. Vol. 4, No. 2. 22-35.
[20]   Kim, H. J. (2010). Investigating the Construct Validity of a Speaking Performance Test. June 24, 2010, Volume 8: 1–30. www.lsa.umich.edu/eli/research/spaan.
[21]   Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. Test validity (pp. 33-45). Hillsdale, NJ: Erbaum.
[22]   Mislevy, R. J. Steinberg, L. S. & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3–62.
[23]   Moshman, R. (2011). Conventional Testing Versus Authentic Assessment in Education. *www.rachaelmoshman.com. Journal of Language and language testing Studies, 10*(2), 120-133; 2011.
[24]   Pishghadam, R. & Shams, M.A. (2013). A New Look into the Construct Validity of the IELTS Speaking Module. *The Journal of Teaching Language Skills (JTLS),*5 (1), Spring 2013, Ser. 70/4. ISSN: 2008-8191. pp. 71-90.
[25]   Rahimpur, M. Salimi, A. & Farukhi, F. (2012). The Effect of Intensive and Extensive Focus on Form on EFL Learners' Written Accuracy", *Theory and Practice in Language Studies,* 2012.
[26]   Seedhouse, P. & Harris, A. (2008). Topic development in the IELTS Speaking Test. *IELTS Research Reports* August 12, 2008/Volume 12, 41-56 www.ielts.org.
[27]   Seyed Erfani, SH. (2014). An Investigation on the Attitudes of IELTS and TOEFL iBT Teachers: A Quantitative and Qualitative Washback Study. 4(1): 88-102, July 2, 2014, www.sciencedomain.org.

[28] Shirinzadeh Aghdam,Z. & Farahani, A. (2012). Speaking as an indicator of general proficiency in placement test. *Language Teaching Research,* 2, 203-229. DOI: 10.1177/13621688980020033.

[29] Stoynoff, S. (2009). Recent Developments in Language Assessment and the Case of Four Large-Scale Tests Of ESOL Ability. *Lang. Teach.* (2009), 42:1, 1–40.

[30] Widdowson, H. (1983). Learning purpose and language use. Oxford, England: Oxford University Press.

[31] Zahedi, K. & Shamsaee, S. (2012). Viability of Construct Validity of The Speaking Modules of International Language Examinations (IELTS vs. TOEFL iBT): Evidence from Iranian Test-Takers. Aug2012, Vol. 24 Issue 3, p263-271.

Elham Zahedkazemi holds an M.A. in TEFL (Teaching English as a Foreign Language) from Ahvaz Science and Research Branch University, Iran. She was born in Dezful. She has had experiencing in area of teaching English for more than 9 years. Ms Elham Zahedkazemi has not been the member of any professional societies yet.