The Journal of Teaching Language Skills (JTLS)
5 (1), Spring 2013, Ser. 70/4
ISSN: 2008-8191. pp. 71-90

A New Look into the Construct Validity of the IELTS Speaking Module

R. Pishghadam *
Associate Professor, TEFL
Ferdowsi University of Mashhad
email: pishghadam@um.ac.ir

M.A. Shams
Ph.D., TEFL
Graduated from Ferdowsi University of Mashhad
email: neoshams@yahoo.com

Abstract

The aim of this study was to investigate the role of linguistic and intelligence factors in the Iranian IELTS candidates' speaking performance. Linguistic factors include depth and breadth of vocabulary knowledge as well as grammar knowledge. Narrative and verbal intelligences represent the non-linguistic factors. The participants included 329 learners who took 5 validated tests and also participated in a simulated IELTS interview session. Model 1 (excluding the intelligence factors) represents the conventional view, whereas Model 2 (including all factors) is proposed for the first time in this study. The Structure Equation Modeling (SEM) was used to analyze the data. Using the SEM, both proposed models were examined to see which one fits the data more. The results of the comparisons made between the parameter estimates and fit indices of the two models demonstrate that Model 2 outfits Model 1, implying that in contrast to the conventional view, intelligence factors do play a significant and undeniable role in developing the speaking construct. Finally, the applications of the findings to promote the construct validity of IELTS are discussed.

Keywords: speaking, IELTS, construct validity, verbal intelligence, narrative intelligence, structural equation modeling

1. Introduction

Achieving an acceptable level of speaking proficiency is a central goal for many language learners. Regardless of their conception of acceptability, a great number of language teachers, Second Language Acquisition (SLA) researchers, and applied linguists are deeply interested in understanding the

Received: 9/30/2012 Accepted: 3/10/2013
* Corresponding author

cognitive, neural, and behavioral correlates of L2 speech fluency and devising ways for improving language learners' oral performance. In spite of numerous reports of remarkable progress made in developing and validating more valid and reliable speaking tests (e.g. Chen & Zechner, 2011; Develle, 2008; Galaczi, 2005, 2010; Huong, 2001; Kim, 2010; Read, 2005; Taylor, 2003), a number of central problems are yet to be solved. According to many scholars in the field (e.g., Brown & Taylor, 2006; Esquinca, Yaden, & Rueda, 2005; Hubbard, Gilbert, & Pidcock, 2006), many of these problems are associated with the rating criteria used in speaking exams. Among various summative and formative speaking tests, high-stakes standardized exams such as IELTS and their rating criteria have a significant effect on one's conception of the speaking construct. This effect is further extended in IELTS preparation courses in particular, and in language classrooms all around the world in general. Therefore, one of the effective ways to refine the current understanding of second language (L2) speaking ability is to analyze, and if needed, revise the rating criteria of the IELTS speaking module.

Among many possible reasons for this gap between leaners' knowledge and the required performance on the exam, the researchers of this study believe that a very important reason comes from the misconceptions held by language learners and their teachers about the nature of L2 speaking ability and the factors that can affect one's speaking performance. One of these misconceptions is that enhancing one's knowledge of grammar and vocabulary will naturally lead to one's better speaking performance. This is rooted in the overestimation of the role of knowledge of grammar (Cutrone, 2009; Katayama, 2007; Lockley & Farrell, 2011; Nishino, 2008; Saito & Ebsworth, 2004) and vocabulary (Horwitz, Horwitz, & Cope, 1986; Ohata, 2005; Young, 1990) in L2 speaking. It is our belief that that although learners' linguistic knowledge has a central role in developing their speaking ability, such knowledge is by no means enough to create a type of performance which usually receives a high score on the exam. To avoid the misconceptions about the speaking construct, the nonlinguistic factors involved in the process should not be overlooked. One of the major nonlinguistic factors can be the cognitive ones, among which intelligence is of great importance. As Fahim and Pishghadam (2007) have shown, intelligence can contribute to English language learning. That is why we are of the view that models delineating the speaking ability may take intelligence into account. In this study, out of different types of intelligence, verbal, and narrative intelligences were examined because they are deeply related to the nature of speaking and narration. In this study, the main point is that including intelligence factors in a model of speaking will increase the

model's viability. That is, taking one's intelligence as well as linguistic knowledge will lead to better predictions of one's L2 speaking performance. Therefore, the main research questions in this study are:

1. Does a model of L2 speaking include only knowledge of grammar and vocabulary (linguistic factors)?
2. Does a model of L2 speaking include narrative and verbal intelligences (nonlinguistic factors) along with knowledge of grammar and vocabulary (linguistic factors)?

2. Theoretical Framework

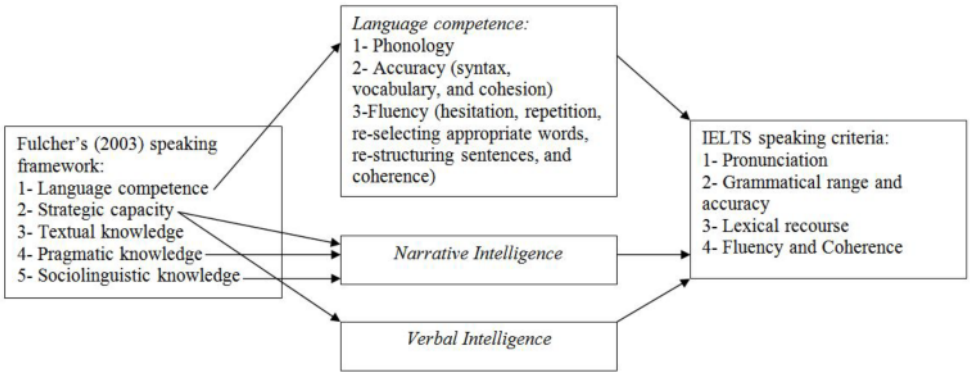
The configuration of the IELTS speaking module reflects the philosophy behind its design. This philosophy can be translated into the theoretical models of language ability, some of which will be discussed later in this paper (see 2.1). A group of research projects focused on the speaking skill move from practice to theory; in other words, they analyze the current status of speaking exams to extract the underlying theoretical assumptions held by test developers (see Chen & Zechner, 2011; Galaczi, 2005, 2010; Kim, 2010; Read, 2005; Shaw, 2003). The current design of the speaking module is the results of several years of study, effort, and several revision projects suggested by IELTS scholars (see Brown & Taylor, 2006; Develle, 2008; Hubbard et al. 2006; Huong, 2001; Taylor, 2001a, 2001b, 2001c, 2003; Taylor & Jones, 2001). The existence, arrangement, and layout of some speaking tasks show the developers' perception of the essential aspects of the speaking proficiency in a foreign language. Here, the theoretical framework for the development of the IELTS speaking module are presented by referring to the research results and revision guidelines reflected in *Research Notes* of ESOL examinations, which can be taken as a valid indicator of the thought patterns and research paradigm adopted by the IELTS team of developers. In addition, to increase the comprehensiveness of this brief review, the other relevant theories of language ability and construct validity are included wherever needed. In the examiner's eye, the speaking construct is delineated via the rating criteria. These criteria are the operationalized format of the main variables involved in speaking. Based on the IELTS criteria for rating the candidates' speaking performance, oral proficiency is reducible to grammar knowledge, vocabulary knowledge (lexical resources), fluency, and pronunciation. Besides, the relevance of the answers to the prompt is labeled as task achievement which also forms a part of the speaking score.

The three parts of the IELTS interview require different sets of speaking abilities. Out of the five types of knowledge proposed by Fulcher

(2003) as the components of speaking ability, it is only the static parts of language competence that can be assumed to manifest almost equally in all the three sections. This is because one's knowledge of phonology, syntax, and vocabulary cannot change over a 15-minute period during the IELTS interview. However, all the other parts of the Flucher's speaking framework refer to the dynamic pieces of knowledge that, under the influence of various factors, may be readily manifested with considerable fluctuation. Strategic competence, textual knowledge, pragmatic knowledge, and sociolinguistic knowledge are dynamic. It is also true for one part of the language competence, namely fluency, which in Fulcher's term, can be observed in hesitations, repetitions, reselection of appropriate words, restructuring sentences, and the coherence of speech provided by the candidate. Among the rating criteria in IELTS, fluency and task achievement represent the dynamic aspects of speech. Compared to the candidates' rather static knowledge of syntax, semantics, and phonology which are relatively easy to observe by the IELTS examiner, fluency has proved to be a much more complex component in the speaking construct. It can be argued that most of the criticisms leveled at the construct validity of speaking exams particularly IELTS are, in one way or another, associated with fluency or other dynamic aspect of speech such as strategic competence, communicative competence, or sociolinguistic knowledge which are eventually observed via fluency. Therefore, it is fair to claim that any attempt to promote the construct validity of the speaking module has to address the issue of fluency.

Intelligence factors can bridge the gap between models of speaking and the IELTS speaking module. Language competence is not the only cognitive trait tested in IELTS. Based on Fulcher's (2003) speaking framework, a range of language-free cognitive abilities are also put to test in any speaking exam. The contribution of these general cognitive abilities to the fluency of speech depends on the structure and procedure of the test. As Luoma (2004) maintains, before discussing the speaking construct, the specification of the exam must be analyzed. According to Segalowitz (2010), appropriateness of communication is specifically relevant to L2 fluency. In Fulcher's (2003) framework, this is addressed by pragmatic and sociolinguistic knowledge. Hymes (1967) considers social appropriateness as a necessary aspect of any communicative act, specially speech. The implicit role of nonlinguistic cognitive factors in L2 speech can be explicitly analyzed by introducing measures of intelligence into models of speaking ability. Figure 1 shows how narrative and verbal intelligences can be used to link Fulcher's (2003) speaking framework to the speaking construct as it is defined in IELTS:

Figure 1. The links between Fulcher's (2003) speaking framework and the IELTS speaking construct.



Verbal intelligence, in simple terms, means the ability to verbalize one's thoughts (Bandura, 1982). A test of verbal intelligence examines the testees' ability to explain concepts in their L1 (Wechsler, 1981, 1997). Narrative intelligence refers to one's ability to perceive and reproduce narrative patterns (Randall, 1999). It can reflect one's pragmatic and sociolinguistic knowledge as well as one's strategic capacity. As it is defined by Randall (1999), narrative intelligence comprises five subabilities namely emplotment, characterization, narration, genre-ation, and thematization. According to him, narrative intelligence is a hybrid of verbal, interpersonal, and intrapersonal intelligences proposed by Gardner (1983). Emplotment refers to the ability of creating and maintaining a central line for narrative discourse; this is done by the use of logical connectors and lexical linkers. The internal dynamics of emplotment resemble some of the skills required to produce a cohesive and coherent oral discourse. Characterization is divided into two main categories: the narrator's understating of the characters in the story, and of her own character as the narrator. These two aspects are reflected in the referencing and descriptive strategies used by the narrator. The third subability, narration, carries the main essence of narrative intelligence. It refers to one's ability to put events, characters, and concepts in narrative patterns and present them with enough productive and perceptive fluency. Genre-ation, the label of the fourth subability, is actually a term coined by Randall (1999). It includes narrator's conception of the general patterns and the general mood of his or her discourse. Maintaining a certain mood (tragic, realistic) shows the narrator's understanding of the minor and major moves taken throughout the story. Finally, thematization refers to one's ability to identify and focus on certain

signs and points in the narrative discourse to emphasize on a theme or attract the audience's attention to specific aspects of the events in the story. The fourth and fifth subabilities together can effectively contribute to the communicative power of the story or the speech produced by the narrator or the IELTS candidate.

3. Method

3.1 Participants

The participants included 329 Iranian language learners with the age range of 18 to 32. Two hundred and fifty eight of them were university students majoring in basic sciences, engineering, and English language and literature and 71 participants were learning English in private language institutes. The samples were collected from four cities of Iran, namely Tehran, Mashhad, Lahijan, and Kashan. The sample included 136 males and 193 females. The personal information revealed by the participants remained confidential during and after the study.

3.2 Instrumentation

Six tests were used to measure the participants' knowledge of grammar, depth and breadth of vocabulary knowledge, verbal intelligence, narrative intelligence, and speaking ability.

The participants' grammar knowledge was measured using the structure module of the TOEFL PBT (ETS, 2005b). The test was taken from an actual exam administered by the Educational Testing Service (ETS); therefore, the validity of the scale was assured. The test included 40 items. Fifteen items present a sentence with one part replaced by a blank. In the next 25 items, each sentence has four underlined words or phrases, out of which the participants had to identify the wrong ones. The Cronbach's alpha was 0.80.

The Depth of Vocabulary scale was used to examine the depth of the participants' vocabulary knowledge (Quin, 1999). The test included 40 items. Each item had a stimulus word and eight choices in two parts. The first four choices (A-D) are in one part and the second four choices (E-H) are in another part. Part 1 includes possible synonyms for the word in question, whereas the other box includes possible collocations of it. Because four choices were allowed altogether and at least one item was chosen from each part, the number of the possible choices in each part ranges from 1 to 3. According to Qian (1999), the reliability of this test was 0.91. The Cronbach's alpha was 0.76 in this study.

The breadth of participants' vocabulary knowledge was examined using the second version of the Vocabulary Levels Test (Schmitt, Schmitt, & Calpham, 2001). The measure is composed of five frequency levels (2,000,

3,000, 5,000, academic, 10,000) called the levels test. The first two levels (2,000 and 3,000) are composed of high frequency words. The 5,000 level is considered a boundary level, and the next two levels consist of words that generally appear in university texts (academic) and low frequency words (10,000). According to Schmitt, Schmitt, and Calpham (2001), the Rasch ability estimates showing the validity of the five sections of this test are as follows: 42.5 (2,000 words level), 45.9 (3,000 words level), 51.0 (5,000 words level), 55.2 (Academic words level), and 61.7 (10,000 words level). The reliability of the different levels of this test was reported as follows; 2,000 (.92); 3,000 (.92); 5,000 (.92); academic (.92); and 10,000 (.96) (Schmitt et. al, 2001). The Cronbach alpha in this study was 0.81.

The verbal scale of the Wechsler's Adult Intelligence Scale (1981) was used to measure the verbal intelligence of the participants. The Farsi version of the WAIS vocabulary subsection developed by Azmoon Padid Institute (1993) in Tehran, Iran, included 40 items. Each item was an isolated word in Persian, and the participants had to define it orally. Each of their answers could be scored in a range of 0 to 2. The maximum possible score was 80. The Alpha Cronbach for the vocabulary subsection in the present study was 0.68. The reliability coefficient (internal consistency) for the verbal IQ was .97. The vocabulary subtest had quite significant correlations (.91-.95) with the verbal scale of the WAIS-III whose concurrent validity was, in turn, established based on high correlation with other valid intelligence scales. For example, Silva (2008) reports the correlations between this test and the Stanford-Binet Intelligence Scale composite scores to be ranging from 0.78 to 0.89.

The test used in this study to measure the participant's narrative intelligence was developed by Pishghadam, Baghaei, Shams, and Shamsaee (2011). They substantiated the validity of this scale via Rasch analysis. This scale includes 23 items and assesses participants' performance on several dynamics of narrative intelligence (Randall, 1999). The scale includes five subsections parallel to the five subabilities of narrative intelligence as defined by Randall: Emplotment, characterization, narration, genre-ation, and thematization. The participants' ability in each item was rated on a scale of 1 to 5, and the total score represented their overall narrative intelligence. According to Pishghadam et al. (2011), the reliability (internal consistency) of this measure is 0.72. The interrater reliability of the scale was 0.83. The tests' Alpha Cronbach was 0.85.

The main instrument used to measure the speaking ability of the participants was the actual specimen of the IELTS speaking module provided by ETS (2005a). The test includes three sections and took between 11 and 15 minutes to administer. In Sections 1 and 3, questions about

familiar and abstract topics were asked, respectively. The question for Section 2 was written on a card and given to the candidates. In all sections, the participants were required to provide oral answers, and their performance in four areas was rated based on the IELTS criteria: pronunciation, grammar, lexical resource, and fluency (coherence). Each participant could obtain a score from 1 to 9 (with half bands) in each of the areas, and the average score was taken as a measure of their overall speaking ability. The validity of this test had already been assured by ETS. The interrater reliability was 0.82.

3.3 Procedure

The data collection phase comprised the administration of the six tests; this phase started in July, 2010 and ended in July, 2011. During this period, the samples were gathered across the four cities used for data collection. Other than the narrative intelligence test which was administered via a movie session and recording the participants' voice and the speaking test which was administered via interview sessions, the other four tests were given to them in traditional setting of paper and pencil exams. At the first phase of the study, the participants took the speaking test, and their performance was rated based on the IELTS scoring criteria. This produced a set of speaking scores on a scale of 1 to 9 with half-band scores. Then, the test of grammar was taken by participants, and each person received a score out of 40. In the next step, the depth of vocabulary test was administered, and the participants were asked to mark four choices altogether for each item. This test produced a set of scores ranging from 0 to 100. Then, the depth of vocabulary test was given to the participants. The participants' scores on this test were given on a scale of 0 to 160. After that the verbal intelligence test was administered during which each participant was presented with one word at a time and asked to explain each word's meaning verbally. The examiner rated the responses with a 0, 1, or 2 depending on how well the participant defined the word. Therefore, the scores can range from 0 to 80 (Wechsler, 1997). The last phase was the administration of the narrative intelligence test. The participants watched the first 10 minutes of a movie (*Defiance*), and then were asked to recount the story. They were also asked to tell their story of the first day of elementary school. The two narratives produced by each participant were then rated by two raters using the Narrative Intelligence Scale (NIS). The average score for the five subabilities of narrative intelligence in the above narrative tasks was taken as the participants' narrative intelligence score.

First of all, the internal reliability of the tests used in the study was calculated using the Cronbach's alpha. After ensuring the reliability of the

scores, all the data were imported into SPSS 18.0 and linked to AMOS 16.0 to be analyzed through SEM. The observed variables in the models represent the collected data and the latent variables represent the hypothetical constructs which are assumed to play a role in developing learners' speaking ability. Two models—one including only linguistic factors and the other one including intelligence factors as well—were linked to the data and their fit indices and parameter estimates were calculated by AMOS.

4. Results

Table 1 shows the descriptive statistics of all the tests administered in the present study. As seen, the scores obtained in the language exams are more diverse, whereas most of the intelligence scores are closer to the mean. In other words, less heterogeneity is observed in the intelligence scores:

Table 1. The descriptive statistics of the six tests administered in the study

	Mean	Std. Deviation	Std. Error of Measurement	Min.	Max.
Grammar	57.11	16.15	0.89	23	98
Depth of Vocabulary	40.93	14.10	0.79	7	88
Breadth of Vocabulary	44.86	19.00	1.04	12	100
Verbal Intelligence	73.18	6.93	0.38	54	93
Narrative Intelligence	55.96	9.87	0.54	36	90
Speaking	54.07	10.83	0.60	33	94

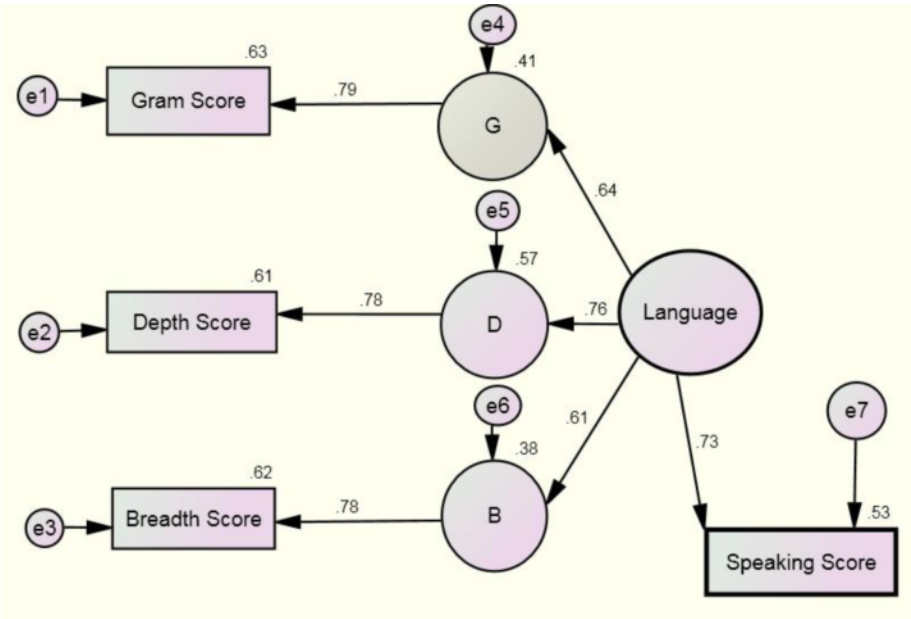
The number of the scores for all the variables was 329. It should be mentioned that each test had its own rating scale (as elaborated in the Method section); however, all the scales here in the table are converted to a scale of 0 to 100 for the sake of easier comparison.

4.1 SEM parameter estimates

Parameter estimates in a SEM model are of two types: 1) The numbers attached to the arrows represent the role each variable plays in determining the variance in the target variable, and 2) the number attached to rectangles or circles (observed or unobserved variables) represent the variance of the target variable which can be explained based on other variables. These are similar to the regression coefficients and explained variances in a regression model; the difference here is that unobserved variables are also included in the model. These unobserved variables, in turn, may be of two types: main variables and error variables. Main unobserved variables represent the factors which might have a role in the model based on the theory but, unlike observed variables, cannot be observed directly. Error variables represent the unexplained variance observed in each variable which could be simply

created due to error in measurement or due to weak modeling in which the role of some latent variables is neglected. Figure 2 shows the SEM model which is based on the mainstream theories for the components of the speaking ability.

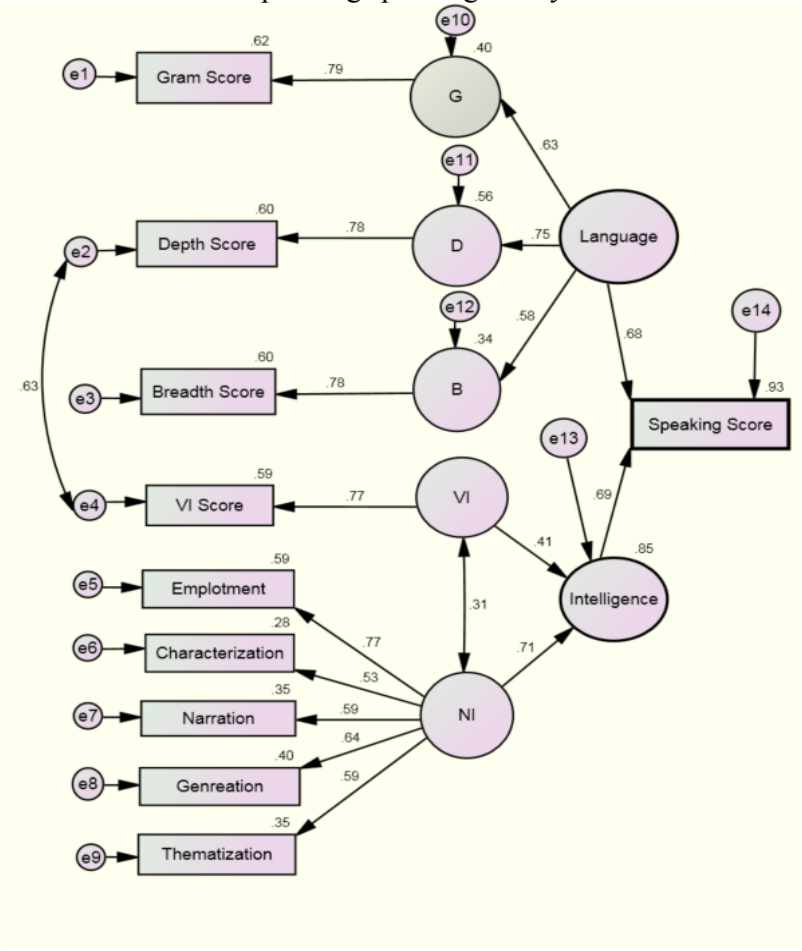
Figure 2. Model 1 with three linguistic factors for explaining speaking ability



According to the parameter estimates shown in Model 1, only 53% of the variance observed in the speaking scores can be justified based on the linguistic competence (labeled "language" in the model). This general competence leads to three other unobserved variables, namely knowledge of grammar, depth of the knowledge of vocabulary, and breadth of the knowledge of vocabulary, which can explain their variance by 41%, 57%, and 38%, respectively. Each of these unobserved variables, in turn, can explain 63%, 61%, and 62% of the variance observed in the actual scores observed in the study. The main question raised over this model is that if 47% of the variance observed in the speaking scores cannot be explained by referring to one's grammar and vocabulary knowledge, then how should it be explained? What other variables could play such a significant role in determining one's speaking performance? One possible answer could be that there are no other factors with a systematic role in the variance of the target variable, and the unexplained variance is totally due to error in

measurement. However, a more reasonable answer is that it is the weakness of the abovementioned theory that has led to such low explained variance for the speaking variance. One way to improve the model is adding other possibly relevant variables which may have an effect on one's L2 speech. We believe that the role of intelligence factors should not be neglected. Therefore, we have devised another model by incorporating two relevant intelligences, namely narrative intelligence and verbal intelligences. If these two variables have a meaningful role in L2 speech performance, then the parameter estimates of this new model must be better than Model 1. Figure 3 presents SEM Model 2 in which both linguistic and nonlinguistic (intelligence) factors are used to model the factors affecting L2 speaking ability:

Figure 3. Model 2 with two intelligence factors and three linguistic factors for explaining speaking ability



As seen in Figure 3, the parameter estimates in Model 2 outdo those of Model 1. This implies that by adding the intelligence factors to the model, the unexplained variance observed in the speaking scores reduces by 40% which is a considerably significant difference compared to Model 1. This means that intelligence factors do have a meaningful role in determining one's speaking ability. But this is not the whole story. A further question can be raised: Do verbal and narrative intelligence have an equal role in one's speaking performance? According to the parameter estimates, the answer is negative. Verbal intelligence and narrative intelligence together can explain 85% of the variance calculated for the general intelligence factor, but their contribution to this variance is not the same. The narrative intelligence parameter is more than the verbal intelligence parameter by 0.30 units. There is also another difference between these two intelligence factors in Model 2. Verbal intelligence is observed via only one observed variable, namely verbal intelligence score, whereas narrative intelligence is observed through five distinct observed variables which are labeled with the five subabilities of narrative intelligence as defined by Randall (1999). The only meaningful correlation between the error variables in this model is found between the e2 and e4—the variables representing the error attached to the measurement of the depth of vocabulary and verbal intelligence, respectively. One possible justification for the relatively high correlation between these two variables (0.63) is that the same factor that may affect the measurement of verbal intelligence has a significant role in affecting the measurement of the depth of vocabulary as well. Given some similarities in the two tests administered for rating these abilities (e.g., relying on one's knowledge of collocation), such significant correlation between the error variables is not surprising.

4.2 SEM fitting indices

SEM fitting indices show SEM models' viability for explaining the variance observed in the data. Table 2 shows the fitting indices of Model 1 (including only language factors) and Model 2 (including both language and intelligence factors). There are two ways to evaluate the indices: 1) comparing the fit indices of each model with the cut-off value, and 2) comparing the fit indices of the two models against each other. The first way shows to what extent the model can be viable, and the second way can be used to decide which model is more appropriate to explain the collected data sets. The term cut-off value refers to the minimum or maximum values suggested for each of the SEM indices. There are several references in mathematics and applied linguistics to decide about the cut-off values, and there are some differences between these references. The main reference to decide about the cut-off values used in this study was adopted from Kaplan

(2009). Among others, six indices are chosen to be presented in Table 2. Each of these indices carries specific information about the models. X^2 is the result of a chi-square test and shows if the data perfectly fits the data. AGFI (Adjusted Goodness of Fit Index) is calculated by a formula that takes into account *df* (Arbuckle, 2007). According to him, IFI (Incremental Fit Index) compares the models' *df* and discrepancy to the baseline model. TLI (Tucker-Lewis Index) depends on the correlation among the variables in the model; it is a good index to compare competing models (Fornell & Larcker, 1981). CFI (Comparative Fit Index) is similar to TLI, but there is a difference: It considers the increment in noncentrality. (Schmacker & Lomax, 2004). Finally, RMSEA (Root Mean Square of Approximation) shows the badness of fit. The lower it is, the better the model fits the data (Schmacker & Lomax, 2004).

Table 2. Fitting indices for model 1 (excluding intelligence) and model 2 (including intelligence)

Fit Index	χ^2/df	AGFI	IFI	TLI	CFI	RMSEA
Acceptable Range	< 3	> 90				< 0.08
Model 1	7.94	0.88	0.89	0.79	0.89	0.14
Model 2	2.86	0.91	0.92	0.90	0.92	0.07

According to the fit indices shown in Table 2, one can see that Model 1 neither fits the data in an acceptable range nor does it outfit Model 2. The most important indices to analyze Model 1 are X^2/df and RMSEA. Because the former shows that the primary conditions for calculating other indices are not available, the latter signal a high and unacceptable level of error in the predictions made by the model. The other four indices are also important in that all show the unfitting nature of Model 1 from different angles. In contract, Model 2 fits the data with acceptable indices. Of course, some of the indices for Model 2 are very close to the cut-off values. For example, TLI is 0.90 which is relatively low value. However, one should pay attention that this index is calculated based on the correlations in the model. While the number of correlations is way smaller than the number of regression lines, the low value of this index can be ignored. What is important is that, all in all, Model 2 both outfits Model 1 and produces acceptable fitting indices. This along with the result of the comparisons made between the parameter estimates of these two models show that including intelligence factors in a model of speaking ability will increase the viability and fitness of the model. It also shows the validity of the theory based on which such improved model was devised.

5. Discussion

The present study was launched to test a central point according to which intelligence factors have a significant role in determining one's L2 speaking ability. To this end, two SEM models were proposed. Model 1 reflected the mainstream theory about the factors affecting one's L2 speech performance. In this model, three linguistic variables, namely knowledge of grammar, depth of vocabulary knowledge, and breadth of vocabulary knowledge were used to predict the speaking scores. Model 2 includes both linguistic and nonlinguistic (intelligence) factors. The results of the comparisons made between the parameter estimates and fit indices of the two models show that Model 2 fits the data better. Therefore, it can be argued that including intelligence factors in a model of speaking is a reasonable suggestion supported by substantial statistical evidence. The researchers believe such results should work as a ground platform to raise deeper questions and to face more challenging research problems proposed with an explanatory approach. The details of the interrelationship between linguistic and intelligence factors must be further studied.

With respect to verbal intelligence, the outcomes of this study are compatible with the findings of Fahim and Pishghadam (2007), based on which they have claimed that verbal intelligence contributes to language learning in general. In the same vein, the findings of this study support those of Pishghadam (2009) in which it has been found that verbal intelligence impacts the writing ability of the learners. One possible line of explanation for this finding can be that the speaking ability directly deals with the knowledge of words and grammar. In fact, verbal intelligence is the ability to articulate one's knowledge of the concept in a given language rather than the knowledge itself. The results of the present study revealed that including verbal intelligence in a model of speaking ability can promote the model and produce better fit indices. It implies that one has to resort to a theory of speaking proficiency that accommodates a set of language-free factors that can operate in both languages. All in all, these findings signal the importance of verbal intelligence in developing productive language skills.

Regarding the narrative intelligence, the findings of this study espouse those of Pishghadam and shams (in press) in which they have shown that narrative intelligence is influential in the writing ability. Like the writing ability, speaking is productive, dealing with the narrative power of individuals. This type of association between narrative intelligence and the speaking ability highlights the critical role of cognitive factors in designing speaking proficiency models. Evidently, speaking requires more than the linguistic abilities. Moreover, the role of narrative intelligence in L2

speaking is not as conceivable as the role of verbal intelligence. This is due to two main reasons: 1) Narrative intelligence, as it is defined by Randall (1999) and operationalized by Pishghadam et al. (2011), may be misunderstood. 2) The narrative nature of some of the mini-tasks included in the IELTS interview may not be evident. Therefore, explaining about the possible links between the dynamics of narrative intelligence and the narrative aspects of the IELTS speaking tasks can shed more light on the significant role of narrative intelligence in a model of speaking as it is found in the present study. According to Randall (1999), successful emplotment is based on distinguishing between the main plot and subplots, perceiving situations as discrete temporal units with beginnings, middles, and ends, maintaining central storylines and linking events in a consequential order. As one can see, all of these abilities contribute to one's pragmatic knowledge. For example, in Section 1 of the IELTS interview, when candidates have to provide a very brief introduction to their current situation, such abilities can be effectively helpful to produce a coherent piece of speech. For instance, if a candidate fails to maintain the central storyline and distinguish between the plot and subplot while providing a very brief story of his or her life, they will digress and lose mark on IELTS due to failure in task achievement, which is one of the rating criteria. Pragmatic knowledge refers to the appropriateness of the use of linguistic devices in real contexts. When the candidates express their compassion for the characters that appear in the mini-stories during the interview (characterization), imagine a dramatic shape for the events (genre-ation), or explain about the message embedded in their narrative discourse (thematization), they are actually showing their pragmatic knowledge to manage the flow of speech in the desired direction.

Sociolinguistic knowledge of the candidates is mainly manifested through characterization and thematization. "Imagining the characters' thoughts and feelings" (Randall, 1999, p. 20) is one of the mental abilities that can show one's knowledge of the social aspects of language which are mainly realized in the relationship between the characters. For example, when in Section 3 the candidates are asked to support their opinion on a certain issue and compare it to the opinions expressed by other social groups, they have to be able to see the issue from alternative perspectives; therefore, imagining other characters' thoughts becomes a valuable ability that can help the candidates to improve the sociolinguistic aspect of their speech. In addition, they can mention and analyze recurrent patterns of events (thematization) to make points and convey their intended meaning. This will improve their communicative capacity which is another dimension of one's sociolinguistic knowledge.

Strategic capacity of the IELTS interviewees is manifested through narration and genre-ation. According to Randall (1999), a good narrator should use neither too much detail nor too little. He also emphasizes the speakers' ability to sustain the interest of the audience by using appropriate rhetoric moves as an important part of narration. The strategic moves made by the IELTS candidates are closely related to the mood speech. According to Randall (1999), a good narrator should be able to maintain a particular tone (e.g., realist) in his her speech. This is very important in the IELTS interview, especially Section 3 in which the candidates maintain a certain line of argument and produce a coherent answer. In fact, Fulcher's (2003) strategic capacity, which contributes to the coherence of speech in IELTS, shows one's ability to narrative and genre-ate (Randall, 1999). All of the above explanations along with the significant statistical findings of this study signal the need for changing the theories of speaking proficiency and altering pedagogical practices for teaching this skill in a way that is respectful to the place of intelligence factors in language learning and testing.

The results of this study can be effectively applied in the teaching and testing of L2 speaking. By deigning pedagogical tasks whose aim is to improve L2 learners' verbal and narrative intelligences, English teachers and material developers can help learners to enhance their speaking skills. For example, devising a set of exercises similar to the items included in the test of verbal intelligence in an L2 can prepare IELTS candidates for the speaking module and increase their chance for obtaining higher scores. A similar strategy can be applied to make the candidates familiar with the dynamics of narrative intelligence and using this new knowledge to enhance their communicative capacity in conversational narrative situations such as IELTS interview. The literature of speaking research shows that there are a few instances of paying attention to the narrative competence of L2 learners. The clarity and systematic potentials of Randall's (1999) framework to identify and develop the subabilities of narrative intelligence can be effectively used to improve narrative intervention programs which are launched with linguistic purposes (e.g., Dobson, 2005; Hussein, 2008). The result of this study can also contribute to the construct validity of speaking test. For example, the rating criteria of IELTS speaking module can be revised by taking the narrative aspects of the tasks into account. A project was conducted by Ball (2000) and Ball and Willson (2002) to investigate the story-telling tasks included in YLE exam. IELTS researchers can follow that line of inquiry to investigate the narrative aspects of IELTS speaking module. This could also have an effect on the criteria and procedure for selecting or designing the speaking tasks. Validating speaking tasks is

considered as a crucially important phase of any revision project (O'Sullivan, Weir, & Saville, 2002; Saville, & O'Sullivan, 2000). In addition, any changes made in the speaking tests will affect the teaching methods through washback after a while (Esquinca et al., 2005).

The findings of this study generate a set of new research questions which can be addressed in the future. The role of intelligence factors in receptive skills i.e., reading and listening can be analyzed using a similar SEM approach. More diverse samples form populations with different L1 backgrounds can be included in the future projects to exclude any possible extraneous variables created in the Iranian context. This study was totally based on behavioral correlates of speech fluency. A more comprehensive research project would include neural correlates as well. This study focused only on the speaking module of the IELTS exam; future studies can be conducted using other standardized or validated tests of speaking to produce a more representative set of research results that can reflect the current situation of the speaking assessment in general.

References

Arbuckle, J. A. (2007). *Amos 16.0 user's guide.* Chicago: AMOS Development Corporation.

Ball, F., & Wilson, J. (2002). Research projects related to YLE speaking tests. *Research Notes*, *7*, 8-10.

Bandura, A. (1982). Self efficacy mechanism in human agency. *American Psychologist, 37*, 122-147.

Brown, A., & Taylor, L. (2006). A worldwide survey of examiners' views and experience of the revised IELTS speaking test. *Research Notes*, *26*, 14-18.

Chen, M., & Zechner, K. (2011). Computing and evaluating syntactic complexity features for automated scoring of spontaneous nonnative speech. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 722-731.

Cutrone, P. (2009). Overcoming Japanese EFL learners' fear of speaking. *University of Reading Language Studies Working Papers, 1*, 55-63.

Develle, S. (2008). The revised IELTS pronunciation scale. *Research Notes*, *34*, 36-39.

Dobson, S. (2005). Narrative competence and the enhancement of literacy: Some theoretical reflections. *International Journal of Media, Technology, and Lifelong Learning, 1*(2), 1-14.

Educational Testing Service (2005a). *IELTS specimen material.* Cambridge: Cambridge University Press.

Educational Testing Service. (2005b). *TOEFL actual tests: 7 full-length paper-based exams administered in the past by ETS.* Cambridge: Cambridge University Press.

Esquinca, A., Yaden, D., & Rueda, R. (2005). Current language proficiency tests and their implications for preschool English language learners. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.). *Proceedings of the 4th international symposium on bilingualism* (pp. 674-680). Somerville, MA: Cascadilla Press.

Fahim, M., & Pishghadam, R. (2007). On the role of emotional, psychometric, and verbal intelligences in the academic achievement of university students majoring in English language. *Asian EFL Journal, 9*, 240-253.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 48*, 39-50.

Fulcher, G. (2003). *Testing second language speaking.* London: Longman.

Galaczi, E. (2005). Upper main suit speaking assessment: Toward an understanding of assessment criteria and oral examiner behavior. *Research Notes*, *20*, 16-19.

Galaczi, E. D. (2010). Peer-peer interaction in a paired speaking test: The case of FCE. *Research Notes*, *42*, 22.

Gardner, H. (1983). *Frames of mind.* New York: Basic books.

Horwitz, E. K., Horwitz, M. B., & Cope, J. A. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125-132.

Hubbard, C., Gilbert, S., & Pidcock, J. (2006). Assessment processes in speaking tests: A pilot verbal protocol study. *Research Notes*, *24*, 14-19.

Huong, T. T. T. (2001). The predictive validity of the International English Language Testing System (IELTS) test. *Post-Script, 2*(1), 66-96.

Hussein, J. W. (2008). An existential approach to engaging adult learners in the process of legitimizing and constructing meanings from their narrative knowledge. *Action Research, 6*, 391-420.

Hymes, D. (1967). Models of interaction of language and social setting. Journal of Social Issues, *33*(2), 8-28.

Kaplan, D. (2009). *Structural equation modeling.* London: Sage Publication.

Katayama, A. (2007). Japanese ESL students' preferences toward correction of classroom oral errors. *The Asian EFL Journal Quarterly, 9*, 289-305.

Kim, H. J. (2010). Investigating the construct validity of a speaking performance test. *Span Fellow Working Papers in Second or Foreign Language Assessment, 8*, 1-30.

Kim, M. (2001). Detecting DIF across the different language groups in a speaking test. *Language Testing, 18*, 89-114.

Kim, Y. (2009). An investigation into native and nonnative teachers' judgments of oral English performance: A mixed methods approach. *Language Testing, 26*, 187-217.

Lockley, T., & Farrell, S. (2011). Is grammar anxiety hindering English speaking in Japanese students? *JALT Journal, 33*(2), 175-190.

Luoma, S. (2004). *Assessing speaking.* Cambridge: Cambridge University Press.

Neville, H., Mills, D., & Lawson, D. (1992). Fractioning language: Different neural subsystems with different sensitive periods. *Cerebral Cortex, 2*, 244-258.

Nishino, T. (2008). Japanese secondary school teachers' beliefs and practices regarding communicative language teaching: An exploratory survey. *JALT Journal, 30*, 27-50.

O'Sullivan, B., Weir, C. J., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing, 19*(1), 33-56.

Ohata, K. (2005). Language anxiety from the teacher's perspective: Interviews with seven experienced ESL/EFL teachers. *Journal of Language and Learning, 3*, 133-155.

Pan, Y., & Pan, Y. (2011). Conducting speaking tests for learners of English as a foreign language. *The International Journal of Educational and Psychological Assessment, 6*(2), 83-100.

Pishghadam, R. (2009). Emotional and verbal intelligences in language learning. *Iranian Journal of Language Studies, 3*(1), 43-64.

Pishghadam, R., Baghaei, P., Shams, M., & Shamsaee, S. (2011). Construction and validation of a narrative intelligence scale with the Rasch rating scale model. *International Journal of Educational and Psychological Assessment, 8*(1), 75-90.

Pishghadam, R., & Shams, M. (in press). Pinpointing the classifiers of English language writing ability: A discriminant function analysis approach. *Journal of Applied Research in English.*

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*, 282-308.

Randall, W. L. (1999). Narrative intelligence and the novelty of our lives. *Journal of Aging Studies, 13*(1), 11-28.

Read, J. (2005). Applying lexical statistics to the IELTS speaking test. *Research Notes*, *20*, 12-16.

Saito, H., & Ebsworth, M. (2004). Seeing English language teaching and learning through the eyes of Japanese EFL and ESL students. *Foreign Language Annals, 37*, 111-124.

Saville, N., & O'Sullivan, B. (2000). Developing observation checklists for speaking-tests. *Research Notes, 3*, 6-9.

Schmitt, N., Schmitt, D., & Calpham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing, 18*, 55-88.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling.* London: Lawrence Erlbaum Association.

Segalowitz, N. (2010). *Cognitive bases of second language fluency.* New York: Routledge.

Shaw, S. D. (2003). CELS speaking assessment: Toward an understanding of oral examiner and test-taker behavior. *Research Notes*, *13*, 16-19.

Silva, M. A. (2008). Development of the WAIS-III: A brief overview, history, and Description. *Graduate Journal of Counseling Psychology, 1*, 117-135.

Taylor, L. (2001a). The paired speaking test format: Recent studies. *Research Notes*, *6*, 15-16.

Taylor, L. (2001b). Revising the IELTS speaking test: Developments in test format and task design. *Research Notes, 5,* 2-4.

Taylor, L. (2001c). Revising the IELTS Speaking Test: Retraining IELTS examiners worldwide. *Research Notes*, *6*, 9-11.

Taylor, L. (2003). The Cambridge approach to speaking assessment. *Research Notes*, *13*, 2-4.

Taylor, L., & Jones, N. (2001). Revising the IELTS speaking test. *Research Notes, 4*, 9-11.

Vidakovich, I., & Galaczi, E. (2009). ILEC speaking: Revising assessment criteria and scales. *Research Notes*, *35*, 29-34.

Wechsler, D. (1981). *Wechsler adult intelligence scale revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *WAIS–III administration and scoring manual*. San Antonio, TX: Psychological Corporation.

Young, D. J. (1992). Language anxiety from the foreign language specialists' perspective: Interview with Krashen, Omaggio Hadley, Terrell, and Rardin. *Foreign Language Annals, 25*, 157-172.