ðə dʒərnəl əv læŋgwədʒ titʃɪŋ ænd lərnɪŋ

J L T L

ðə dʒərnəl əv læŋgwədʒ titʃɪŋ ænd lərnɪŋ

http://www.jltl.org/jltl/

The Journal of Language Teaching and Learning, 2011–1(2), 30-43

Multi-componential Models of L2 Listening

Minhee Eom*

Abstract

This study evaluated four hypothesized L2 listening models to examine underling structures of L2 listening. These listening models represented different underlying structures of thirteen listening ability variables derived from the scores of a listening section of the TOEFL (*N* = 450). Model 1 posited a single factor structure model of general L2 listening; Model 2 a two-factor structure of short and long textual lengths; Model 3 a three-factor structure of a general language attribute and two types of comprehension; and Model 4 a four-factor structure of two types of language knowledge and the same two comprehension factors of Model 3. Using confirmatory factor analyses with LISREL, the study compared the overall model fit indices of Chi-Square ($\chi^2{}_M$) Test, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Root Mean Square Residuals (RMSR) and concluded that Model 4 was the best fit model of listening. In the selected four-factor model, two language factors represented organizational and functional knowledge, whereas the other two suggested varying degrees of comprehension complexities. The selected model empirically supported the framework of cognitive models of language ability and other theoretical descriptions of language knowledge and comprehension processes.

Keywords: L2 listening model, cognitive language model, underlying structures, CFA, overall model fit

1. Introduction

L2 listening is often discussed in terms of taxonomies of skills (Mendelsohn, 1994; Richards, 1983; Rost, 2002). While these taxonomies provide a valuable basis for teaching practices such as developing curriculum instructional objectives, it is rare to find theoretical implications ascribed to these classifications.

Listening is an interactive and cognitive processing of various resources; linguistic knowledge, knowledge of the co-text, contextual knowledge, and general world knowledge interact at all levels of

* The University of Texas-Pan American, USA, e-mail: meom@utpa.edu

listening (Buck, 2001). Such an engagement of various knowledge sources suggests a possible multi-componential aspect of L2 listening and can be understood in conjunction with the framework of a multi-componential language ability model (Bachman, 1990; Bachman & Palmer, 1996). Using confirmatory factor analysis, this study examines theoretical dimensions of various L2 listening abilities and suggests a model for L2 listening. The result may contribute to a meaningful understanding of elements and dimensions of L2 listening.

1.1 Defining Language Ability

In the 1980's, there was vigorous research on a general language proficiency factor (Oller, 1983; Carroll, 1983). Oller (1983) defined language proficiency based on empirical data and proposed 'unitary trait hypothesis' which defined language as an integrated system. However, this claim was discredited because of a possible methodological flaw and the use of a limited language type (academic language only). Carroll (1983) provided a compromising viewpoint between 'unitary ability hypothesis' and the 'divisible competence hypothesis.' He assumed that there is "a general language ability" of different language skills but, at the same time, that language skills can be developed and specialized to different degrees, or at different rates. There was an overall degree of advancement of language skills, but language skills could be also separately recognized and measured. However, Carroll did not refer to the nature of the general factor but hinted that it could best be measured by a sort of oral interview judgments.

The concept of a general factor gave its way to the communicative competence movement. As an incipient step for communicative movement, Hymes (1972) expanded the Chomskyan definition of competence and performance. With respect to the definition of performance, Hymes distinguished actual instances of language use in real time from abstract performance models of underlying knowledge involved in language use. The performance model for 'ability for use' included a sociolinguistic model of language knowledge and a broadly psychological model of language performance, which were subsumed under the term communicative competence. The 'ability for use' represented the underlying potential for performances and distinguished itself from actual instances of use. In a similar vein, Canale and Swain (1980) distinguished grammatical competence from sociolinguistic competence and introduced strategic competence as part of language ability. It was of a compensatory function that language learners could use to prevent a breakdown of communications in the case when grammar and social linguistic competence failed. Canale (1983) made a further distinction between sociolinguistic competence and discourse competence (cohesion and coherence) about how the elements hung together in order to have coherent meaning, such as genre. Their models were purely theoretic and not empirically tested.

The Communicative Language Ability (CLA) model proposed in Bachman (1990) and further refined in Bachman and Palmer (1996) is widely studied in research and provides ground for theoretical constructs of language use (Leung & Lewkowicz, 2006; Phakiti, 2003; 2008; Skehan, 1998). This multi-componential model includes such the factors as language knowledge, strategic competence, topical knowledge, and affect. Among these elements that constitute language ability, language knowledge and strategic competence are critical and mostly relevant to the current study. Language knowledge consists of various types of knowledge. Organizational knowledge controls grammatical aspects of language including knowledge of sound system, vocabulary, syntax, and graphology. On the other hand, functional knowledge governs sociocultural and illocutionary functions of communications. Strategic competence is a set of metacognitive strategies that manage language use and other cognitive activities. Language user's topical knowledge, language knowledge and affective schema are integrated into the metacognitive strategy.

The above mentioned language models are grounded in the cognitive framework where language abilities are defined in terms of mostly intrapersonal, and not much of interpersonal, development. Thus, they are not proper models for interpersonal and sociocultural interactions of language uses. In the mid-nineties, sociocultural movements entered into the field of second language research changing the landscape of L2 research (Firth & Wagner, 1997; Lantolf, 1996; 2000; Lantolf & Appel, 1994). Language ability models are enhanced with more sociocultural and communicative attributers (Celce-Murcia, Dörnyei, & Thurrell, 1995) and recognize needs for interactional and socio-cognitive approaches (Chalhoub-Deville, 2003; Chalhoub-Deville & Deville, 2005; McNamura 1997; Young, 2000).

Despite the importance of alternative views, this study lies in the realm of cognitive models and examines L2 listening models with respect to language knowledge and cognitive processes. While language ability models delineate the scope and component of language ability in general, it is necessary to consider issues that may be critical to the mode of L2 listening.

1.2 L2 Listening Comprehension

While listening is a complex process of various sources, linguistic knowledge is the one L2 learners have most difficulty with (Buck, 2001). There is no simple definition of linguistic knowledge, but the following have been identified as significant linguistic elements. First, vocabulary knowledge is a linguistic source affecting listening comprehension. The studies of Nissan, Devincenzi and Tang (1995), Freedle and Kostin (1998), and Kostin (2004) found that word frequency and idioms were significant predictors of the item difficulty index in TOEFL listening tests. However, comprehending the word meaning in aural input requires phonological processing in addition to knowledge of word meanings. That is, knowing the word meaning in writing does not guarantee the comprehension of the word in listening. In listening, comprehension of meaning involves parsing individual sound segments and tone unit boundaries. A tone unit refers to a string of words positioned between pauses and has important semantic and syntactic functions (Brown, 1990). It also represents the amount of information a person can comfortably process in short-term memory.

In addition, the knowledge of prosody such as rhythmic patterns and intonation affect listening. English learners should be able to distinguish different ways of marking stress and recognize stress patterns rapidly and accurately to parse the given structure of the message. Learners familiar with explicitly pronounced speeches designed for English instructions may find it quite difficult to understand normal spoken English. Function (grammatical) words such as conjunctions, prepositions and pronouns lose their stress in utterances and become hard to parse (Brown, 1990). Also, the knowledge of paralinguistic cues can help listeners to parse speech. Speakers employ various types of intonation to express their intentions. Listeners need to properly interpret marked intonation patterns to understand speaker's intentions and emotional status (Crystal, 2004). This prosodic awareness is strictly part of phonological knowledge.

Another noticeable part of listening is functional knowledge that controls communicative nature of listening. Numerous researchers, if not all, state that listening comprehension does not consist of understanding what words and sentences mean, but consists of understanding what speakers mean (Buck, 1995; Mendelsohn, 1994; O'Malley, Chamot & Kupper, 1989). For successful communicative listening, listeners need to interpret speakers' intentions properly the same way speakers need to express their intentions properly. It is not decontextualized knowledge like word meanings and prosody, but knowledge of contexts of conversations or texts. Listeners may perfectly understand literal meanings but not necessarily understand messages speakers try to convey.

In addition to linguistic knowledge, listening involves cognitive features of comprehension. Anderson (2004) proposes three stages of language comprehension applicable to both L1 and L2. They include perceptual processing, parsing, and utilization. At each stage language input is decoded, translated to meanings, and stored in long term memory. This classification of comprehension phases provides much insight to the nature of comprehension processing, but the linear progress of three stages with utilization at the end may not be sufficient to explain interactive and complicated nature of L2 listening. However, it may infer a challenging nature of utilization as the final phase of this linear process. While Anderson conceptualizes the comprehension process in a linear manner, Rost (2005) emphasizes the interaction of different stages in processing input. With some apparent similarities, Rost explains L2 listening comprehension in three stages: decoding, comprehension and interpretation. The decoding phase is where listeners recognize lexical items and parse propositions. In the comprehension phase, listeners connect input to relevant knowledge sources. The final phase of Rost's model involves listeners' interpretation with respect to response options. It seems to represent the interactive nature of L2 listening activities. The first two stages of Rost's classification appear to embrace all of the three stages of Anderson's model. In that sense, Rost's model appears comprehensive and engaging with L2 listening activities.

Despite complicated interactions in the three stages of comprehension, it seems that some listening task can be more challenging than others. For example, inference is a significant and challenging part of comprehension processing as it requires mapping pieces of information across a given text and connecting newly processed information with existing knowledge in memory. Listening comprehension engages complex syntheses of existing knowledge and newly incoming information (Buck, 2001), and inferencing is not merely a guess, but hypothesis formations and modifications (Mendelsohn, 1994). Eventually, listeners need to make hypotheses about the overall meaning of the listening passage in the beginning, then to modify them as more information comes along.

2. Methods

2.1. Aim of Study

Based on cognitive theoretical models of language abilities and major attributes of L2 listening comprehension, this study hypothesized four models of L2 listening and evaluated the overall model fits to select a model representing L2 listening abilities. The proposed models varied in terms of the number and characteristics of the latent factors underlying the listening ability variables of this study.

2.2. Instrument

This study used the listening comprehension section of a Paper Based Test (PBT) of TOEFL (ETS, 1995). The listening section consisted of 50 multiple-choice items with four alternatives across three parts. Part A comprised of 30 items and 30 short conversations between two speakers. Each conversation lasted approximately 6 to 10 seconds and was followed by one question about the conversation. Part B contained two longer dialogues that lasted approximately 1 minute 30 seconds. After each conversation, test takers heard four to five questions about the conversation. Part C had three monologue input texts followed by four to five questions. All the aural texts and questions were spoken only once.

2.3. Participants

Approximately 450 Korean learners of English in Korea took the listening section of TOEFL sample test in 2006. Sixty percent of the participants were secondary school students who were enrolled in TOEFL preparation courses in private English institutions. The rest of the participants were university students or adult learners who studied English in Korea for higher education, employment or promotion opportunities.

2.4. Data Analysis: Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) allows researchers to construct models in advance (Bollen, 1989; Joreskog & Sorbom, 1993; Kline, 2005). Researchers may determine the number of latent variables and may have prior hypotheses about the effect of the latent variables on the observed variables in terms of directionality and magnitudes. Unlike CFA, exploratory factor analysis (EFA) examines emerging patterns of underlying structures without pre-defined models. Eom (2010) used EFA and conceptualized two underlying factors of listening as language and comprehension attributes. On the other hand, this current study uses an alternative model situation of CFA in which the researcher specifies several alternative models or competing models and selects a single model based on an analysis of empirical data (Joreskog & Sorbom, 1993).

Confirmatory factor analysis consists of five stages: model specification, model identification, model estimation, model evaluation, and model re-specification (Bollen, 1989; Kline, 2005). Using LISREL, this study had already processed the cycle of these modeling stages. The models used in this study were respecified in a way that allowed covariances of four sets of measurement errors: COV(x_5, $x9$), COV(x_1, $x9$), COV(x_1, $x12$), and COV(x_9, $x12$). The initial models had standard residuals too big to ignore and needed respecification that allowed some errors to covary.

Theoretically, it was reasonable to assume covariance of errors. Listening is a complex process of various sources, and the factors this study identified would not exhaustively account for the whole spectrum of listening abilities. However, with more error covariances, models statistically tended to produce more good fits. Thus, this study maintained a minimum set of error covariance across all the models. With the four covaried sets of errors, the study was able to find a good fit model among the four hypnotized models.

2.4.1. Observed Variables (x): Listening Abilities

This study used thirteen ability variables developed based on listening taxonomies (Richards, 1983; Rost, 2002) with consideration of the test structure. The test items were coded to these listening abilities by two coders. After independent coding, the coding results were compared. For the disagreed items, the coders met in person and reached agreements after discussion. The total score reliability (Cronbach's Alpha) was .89. See Table 1 for the descriptive statistics and the factor associations between the observed variables and the latent variables.

Table 1
Observed Variables and Latent Variables

Observed Variables	No. of items	Mean	M1	M2	M3	M4
x_1: Recognizing word boundaries	6	.682	$\lambda_{1,1}$	$\lambda_{1,1}$	$\lambda_{1,1}$	$\lambda_{1,1}$
x_2: Recognizing homonyms in alternatives	2	.706	$\lambda_{2,1}$	$\lambda_{2,1}$	$\lambda_{2,1}$	$\lambda_{2,1}$
x_3: Processing marked stress patterns of function words	3	.658	$\lambda_{3,1}$	$\lambda_{3,1}$	$\lambda_{3,1}$	$\lambda_{3,1}$
x_4: Processing marked intonation patterns for attitudinal messages	4	.765	$\lambda_{4,1}$	$\lambda_{4,1}$	$\lambda_{4,1}$	$\lambda_{4,1}$
x_5: Interpreting speakers' indirect speech about their feelings, opinions, and status	3	.809	$\lambda_{5,1}$	$\lambda_{5,1}$	$\lambda_{5,1}$	$\lambda_{5,2}$
x_6: Interpreting what speakers want listeners to do	4	.697	$\lambda_{6,1}$	$\lambda_{6,1}$	$\lambda_{6,1}$	$\lambda_{6,2}$
x_7: Processing selective propositions in low density texts	4	.594	$\lambda_{7,1}$	$\lambda_{7,1}$	$\lambda_{7,2}$	$\lambda_{7,3}$
x_8: Making inference in low density texts	4	.567	$\lambda_{8,1}$	$\lambda_{8,2}$	$\lambda_{8,2}$	$\lambda_{8,3}$
x_9: Processing selective propositions in high density texts	5	.726	$\lambda_{9,1}$	$\lambda_{9,2}$	$\lambda_{9,2}$	$\lambda_{9,3}$
x_{10}: Processing scattered propositions high density texts	4	.642	$\lambda_{10,1}$	$\lambda_{10,2}$	$\lambda_{10,3}$	$\lambda_{10,4}$
x_{11}: Making inference on selective propositions in high density texts	3	.601	$\lambda_{11,1}$	$\lambda_{11,2}$	$\lambda_{11,3}$	$\lambda_{11,4}$
x_{12}: Making inference on scattered propositions in high density texts	3	.516	$\lambda_{12,1}$	$\lambda_{12,2}$	$\lambda_{12,3}$	$\lambda_{12,4}$
x_{13}: Processing main ideas/topics in high density texts	5	.700	$\lambda_{13,1}$	$\lambda_{13,2}$	$\lambda_{13,3}$	$\lambda_{13,4}$

2.4.2. Models with Latent Variables (λ)

Model 1 (Null Model) hypothesized unidimensionality of listening ability and postulated a single underlying factor. It was analogous to the unitary trait hypothesis of Oller (1983).

Model 2 hypothesized two underlying factors assuming that input text lengths were a critical aspect underlying listening abilities. The observed variables of short texts were factored on the first factor and those of long texts were on the second factor. This model was different from the two factor model found via exploratory factor analyses in Eom (2010).

Model 3 hypothesized three underlying factors assuming that language knowledge and comprehension processes represented the dimensions of listening abilities. In the previous study of Eom (2010), the two factors of L2 listening were conceptualized as language and comprehension. Model 3 and Model 4 are developed based on these two factors dividing them into sub-factors. That is, keeping language knowledge as a single factor, Model 3 divided comprehension processing into two factors depending on the degree of comprehension complexity. Difficulty of target propositions to interpret and contextual propositions in texts could affect listening. In addition, certain tasks, like making inference, could be more challenging.

Model 4 hypothesized four underlying factors of listening abilities introducing two factors related to types of language knowledge. That is, Model 4 kept the same structure of the comprehension factors of Model 3 and divided the language factor into two separate factors. As the CLA model classifies language knowledge into subcategories (Bachman & Palmer, 1996), Model 4 assumed two factors for language knowledge of listening: phonological and illocutionary knowledge. These knowledge types may represent organizational and functional knowledge in the CLA model.

2.4.3. Model Evaluations

In model evaluations, there are varieties of indices for goodness-of-fit. Kline (2005) expresses concerns with the availability of various fit indices. For example, different articles may use different fit indices, and reviewers may request the inclusion of certain indices they are familiar with. Moreover, it is possible to make selective reporting of favorable fit indices. Coping with these concerns, Kline suggests a minimal set of fit indices that should be reported in the results of SEM analyses: (1) the model Chi Square, (2) the Root Mean Square Error of Approximation (RMSEA), (3) the Comparative Fit Index (CFI), and (4) the Standardized Root Mean square Residual (SRMR). In a similar manner, Hatcher (1994) recommended the Chi-Square test and the Non-Normed Fit Index (NNFI) and the Comparative Fit Index (CFI). This article focused on the discussion of these fit indices following the guidelines of Kline and Hatcher but also would report other indices for a wider reference.

3. Results

All measures of overall model fit are based on the covariance of the sample and the covariance of the population. The fundamental hypothesis of Structural Equation Modeling is that the population covariance matrix of observed variables is equal to the covariance matrix written as a function of the parameters of the model, which is called implied covariance matrix (Bollen, 1989). Thus, the fit indices estimate the closeness of the sample covariance matrix to the population covariance matrix as function of model parameters. The results of fit evaluations are presented in Table 2.

Table 2. Evaluations of overall model fit

	Model 1	Model 2	Model 3	Model 4
Degrees of Freedom	61	60	58	55
Model Chi-Square				
Minimum Fit Function Chi-Square (χ^2)	151.42 (p = .000)	94.17 (p = .003)	81.37 (p = .023)	73.17 (p = .051)
Normal Theory Weighted Least Squares Chi-Square (χ^2)	166.19 (p = .000)	94.15 (p = .003)	75.75 (p = .036)	71.71 (p = .064)
Root Mean Square Error of Approximation				
RMSEA	.062	.036	.028	.026
90 Percent Confidence Interval for RMSEA	.051;.073	.021;.049	.008;.043	.000;.042
P-Value for Test of Close Fit (RMSEA < .05)	.04	.96	.99	1.00
Comparative Fit Index				
CFI	.98	.99	1.00	1.00
Root Mean Square Residual				
RMR	.054	.041	.036	.034
Standardized RMR (SRMR)	.038	.029	.026	.025
Incremental Fit Indices				
Normed Fit Index (NFI)	.98	.98	.99	.99
Non-Normed Fit Index (NNFI)	.98	.99	.99	1.00
Parsimony Normed Fit Index (PNFI)	.76	.76	.73	.70
Absolute Fit Indices				
Goodness of Fit Index (GFI)	.95	.97	.97	.98
Adjusted Goodness of Fit Index (AGFI)	.92	.95	.96	.96
Parsimony Goodness of Fit Index (PGFI)	.63	.64	.62	.59

3.1. Chi-Square (χ^2_M) Test

Among many indices of model fit, the chi-square test (χ^2_M) statistic is most frequently used to evaluate the overall model fit (Kline, 2005; Hatcher, 1994). A model Chi-square test is, in fact, a badness-of-fit measure. χ^2_M –test assumes the null hypothesis (H_0) that the researcher's model is a perfect fit to the population. A rejection of null hypothesis infers that the model is not a perfect fit. Small chi-squares with large *p* values ($p > .05$) result in a retention of the null hypothesis supporting that the model is a good fit. Thus, researchers expect a *p* value bigger than the critical value in order to retain the null hypothesis and to support a good fit of the specified model.

In the current study, the single factor model, Null Model, produces high chi-square ($\chi^2_M = 151.42$, $p = .00$). The value decreases as more factors are added to the model. Model 3 with three underlying factors ($\chi^2_M = 81.37$, $p = .023$) produce a smaller chi-square than Model 2 with two factors ($\chi^2_M = 94.17$, $p = .003$), but its *p* value is still significant to reject the null hypothesis. Only Model 4 has significantly low chi-square value to retain the null hypothesis, $\chi^2_M = 73.17$, $p = .051$. Thus, Chi-Square model tests support that the four factor model (Model 4) is a good fit in the population and the most favorable model among the specified ones.

One of the weaknesses of Chi-Square tests is its sensitivity to sample size (Kline, 2005; Hatcher, 1994). With large sample size, the test may lead to a rejection of the model even though the model may be a good fit. However, all the models of this study are based on the same data with the same sample size, so this concern is not relevant to the comparisons of the models in this study.

3.2. Root Mean Square Error of Approximation

The Root Mean Square Error of Approximation (RMSEA) is also based on a null hypothesis. Unlike χ^2 tests, however, the null hypothesis of RMSEA assumes that the researcher's model is an approximation of the population, not a perfect fit in the population (Kline, 2005). With the RMSEA, researchers do not reject or retain the null hypothesis but measure the degree of falseness of the null hypothesis. That is, the value of the RMSEA index indicates how false the researcher's model is in approximating the reality. The RMSEA is another 'badness-of-fit' index because the higher the value is, the worse the fit. When the value of index increases, the null hypothesis becomes more and more false. Thus, a model with a smaller value suggests a better approximation of the population, and a zero value indicates the best fit. In general, a RMESA value smaller than .05 represents a close approximate fit, while the value between .05 and .08 indicates a reasonable error of approximation. The RMSEA over .10 suggests a poor fit.

Similar to Chi-Square tests, the model fit improves (or the index value decreases) as additional factors are added to models. Null Model reports RMSEA equal to .062 and suggests that the single factor model has a largest error of approximation among all the models evaluated. The three specified models with multiple factors report approximate fits of the models in the population, RMSEA < .05. However, considering that the RMSEA measures the degree of badness/goodness, the model with a smaller value would be more favorable. The value decreases in a greater degree from Model 2 (RMSEA = .036) to Model 3 (RMSEA = .028) than from Model 3 to Model 4 (RMSEA = .026). This indicates that the three factor model shows a considerable improvement from the two factor model and seems almost as good as the four factor model. However, Model 4 has the smallest value, with a RMESA of .026. Thus, it is inferred that among the proposed models, Model 4 is the best model approximating the population.

Related to the RMESA index, LISREL reports the value of the 90% confidence interval. The interval reflects the degree of uncertainty associated with the RMSEA. The lower boundary of the interval is a cut-off value for a good fit whereas the upper boundary is a cut-off value for a poor fit. In general, the lower boundary value must be less than .05 and the upper value lower than .10 in order to be considered an appropriate fit of the model. The interval of Null model (.051; .073) indicates that it is neither a good fit nor a poor fit. All the models except for the null model show acceptable interval values. However, similar to the other fit indices, the interval of Model 4 (.000; .042) presents the best value for a good fit, which is a zero value of the lower boundary. In sum, both RMESA and the 90% interval values support that Model 4 is the optimum model representing an approximation of the reality.

3.3. Comparative Fit Index

Comparative Fit Index (CFI) measures relative improvements in the fit of a specified model compared with a baseline model or a null model. A baseline/null model assumes zero covariance among observed variables in the population. Due to the assumption of unrelated observed variables, the chi-square index of baseline model tends to be large compared to that of the researcher's model. The comparative fit indices show the extent to which the researcher's model improves from the baseline model. If there is no improvement, there is no reason to prefer the researcher's model. In general, CFI values greater than.90 indicate reasonably good improvement of the researcher's model from the base model.

In this study, this index does not make meaningful evaluations because CFIs of all the models are higher than .98. In fact, Model 2 and Model 3 report a CFI of 1.00. With respect to the high CFI values, Kline (2005) states that high comparative fit values are not very impressive because it is scientifically implausible to assume zero covariance. Moreover, a CFI of 1.0 does not mean a perfect fit, but it only means that the chi-square value of the model is smaller than the degrees of freedom, not that the model is a perfect fit.

3.4. Root Mean Square Residuals

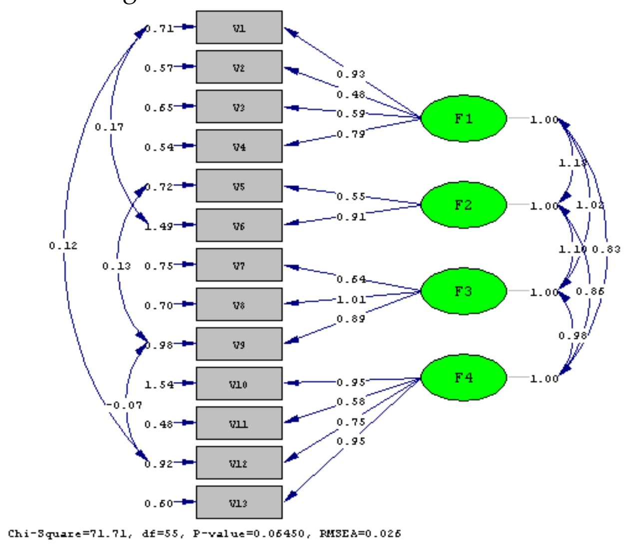
Root Mean Square Residual (RMR) is a measure of the mean absolute value of the covariance residuals. A RMSR that equals .000 indicates a perfect model fit, and the higher the value, the worse the fit. Because RMR is based on unstandardized variables, it is difficult to interpret a given value, unless the observed variables are on the same scale. Thus, Standard Root Mean Square Residual (SRMR) is better for the data with different scales. SRMR transforms unstandardized covariance matrices into standardized correlation matrices to measure the mean absolute correlation residual.

In this study, RMR and SRMR of the models have improved as an additional factor is specified to the model. Null Model with a single underlying factor has the highest values, whereas Model 4 with four factors reports a smallest value, RMR =.034 and SRMR =.025. Although the goodness of fit is not poor for the two factor model (RMR = .041, SRMR = .029), it improves considerably with three factor model (RMR = .036, SRMR = .026). Similar to RMSEA, these residual indices show minuscule decreases from Model 3 to Model 4.

3.5. Evaluation Summary

In conclusion, the overall model fit indices indicate that Model 4 is the best fit among the four models evaluated in this study. See Figure 1 for the factor loadings of Model 4. The three factor model (Model 3) shows a noticeable improvement from the two factor model (Model 2). However, the Chi-Square test supports that Model 4 is apparently better than Model 3. It is important to keep a minimum set of error covariances across the models, because, statistically Model 3 could have been made a good fit if more measurement errors are allowed to covary.

Figure 1
Factor loadings of Model 4



In other words, Model 4 represents the population better than Model 3 allowing the same degree of unknown factors.

4. Discussions

The results of this study support a multi-dimensional nature of L2 listening abilities. The best model of listening abilities has four underlying factors and corresponds to theoretical models of language ability. It is not a surprise to observe similarities between the listening model and language ability models as language ability models are considered overarching models of the sub-skills like listening.

The Communicative Language Ability (CLA) Model postulates two wide-ranging components of language knowledge: organizational knowledge and functional knowledge. The CFA model of this study supports a partition of these two knowledge sources. Factor 1 ($\lambda_{x\ 1}$) of the model represents grammar/phonological knowledge which corresponds to the organizational knowledge, whereas Factor 2 ($\lambda_{x\ 2}$) representing illocutionary knowledge corresponds to the functional knowledge of the CLA model. Organizational knowledge represents grammatical aspects of language including syntax, phonology, textual knowledge, etc. On the other hand, functional knowledge includes several components: ideation functions to allow exchange of information about ideas, knowledge, or feelings; manipulative functions to have other people do things for speakers; heuristic functions to use language for teaching and learning and other cognitive thinking; and imaginative functions for humorous, esthetic and figurative language use. The listening model of this study is mainly limited to manipulative functions. The CLA model provides comprehensive classifications of language knowledge sources. The model serves as a checklist of knowledge types but does not seem to concern relational natures of those knowledge types.

The two language factors bring about considerations in alternative theoretical paradigms. In socio-cognitive or interactional views, Chalhoub-Deville (2003) states that language abilities are considered inextricably enmeshed with a specific context and any inference about the ability becomes questionable. Subsequently, it recommends seeking generalizability through the idea of transfer of conceptual schemes, which allows us to apply knowledge and skills developed in one situation to other situations. Young (2000) also suggests the study of recurring, stable occurrences of contextual interactions. The importance of contexts in defining language abilities, on the other hand, instigates a need for a distinction between transferable knowledge and context-specific knowledge.

It is not yet known what constitutes transferable knowledge, but the factor of phonological knowledge may have transferable attributes. Transferable knowledge can be basic language knowledge insensitive to contexts of its use. The abilities included in the first factor of the listening model in this study are to recognize word boundaries; to recognize homonyms in alternatives; to process marked stress patterns of function words; and to process marked intonation patterns for attitudinal messages. It does not seem unreasonable to say that these abilities are less likely to be affected by the context of communication. On the other hand, the second factor (F2) of this study that supposedly represents functional knowledge includes the abilities to interpret speakers' indirect speech about their feelings, opinions, and status, and to interpret what speakers want listeners to do. These abilities may be more context-sensitive, and, thus, may represent context-specific knowledge. This speculation is not dichotomous. That is, it is not claiming that the transferable knowledge occurs in an absolute absence of contexts, but that different knowledge types may bear a varying degree of contextual attributes. In the extrapolation of the CLA model and socio-cognitive views, the listening model of this study suggests that the organizational knowledge of the CLA model may contain more of the transferable attributes whereas the functional knowledge consists more of the contextual attributes.

In addition to the two language factors, the listening model of this study has identified two factors related to the comprehension process of listening. Anderson (2004) and Rost (2005) propose that comprehension processes may undergo multiple phases of sound decoding, meaning interpretation, and integration of existing knowledge. The two comprehension factors, F3 and F4, of this study speculate a varying degree of comprehension complexity due to the density of propositions to process. As listeners doe not remember the exact wordings of sentences, but remember only propositional meaning representations (Kintsch, 1994; Meyer & McConkie, 1973), the complexity of resulting interpretations may affect the comprehension process.

In the four-factor listening model of this study, F3 ($\lambda_{x\,3}$) represents simple comprehension with a smaller set of propositions involved. It includes the abilities to process selective propositions in low density texts; to make inference in low density texts; and to process selective propositions in high density texts. On the other hand, F4 ($\lambda_{x\,4}$) represents complex comprehension that engages a large set of propositions including the abilities to synthesize scattered propositions in high density texts; to make inference on selective propositions in high density texts; to make inference on scattered propositions in high density texts; and to process main ideas/topics in high density texts. In these definitions of abilities, 'selective propositions' indicate a single or two propositions as target information to comprehend, and 'low density texts and 'high density texts' refer to text length. Thus, the listening model confirms a varying degree of complexity in comprehension depending on the density of target propositions as well as textual propositions.

In addition to the propositional density of target and contextual texts, the complexity of tasks is also to be considered. For example, making inferences about textual information can be more challenging than recalling of simple information as it reaches and uses the last phase of comprehension processes (Anderson, 2004). In a similar vein, Freedle and Kostin (1998) and Kostin (2004) find inference tasks one of the predictors of item difficulty in standardized listening tests. They support two separate factors of comprehension processes in the multi-componential listening model.

When the CLA model is considered, these comprehension factors may correspond to strategic competence. Its functions are to match the new information with existing information and make efficient uses of existing language ability. It may be said that comprehension is a non-language attribute in language use that is realized through strategic competence. Despite its significance, strategic competence is basically defined as uses of meta-cognitive strategies and criticized for its limitation (Phakiti, 2003; 2008). Nonetheless, various types of strategies have been studied for communication (Faerch & Kasper, 1983), second language learning (Oxford, 1990; O'Malley & Chamot, 1990), listening (O'Malley, Chamot & Kupper, 1989) as well as for proficiency in listening (Vandergrift, 2003; 2005). The multi-factor listening model of this study implies that the strategic use may differentiate not only various types but also different degrees of engagements. That is, different listening activities or tasks may call for a simpler or more advanced use of strategic competence.

5. Conclusion, Limitation, and Future Research

This study examines the four listening models hypothesizing different underlying factor of listening abilities: Model 1 hypothesizes a single factor structure; Model 2 a two-factor structure of short and long text length; Model 3 a three-factor structure of a general language factor and two factors of varying comprehension complexity; and Model 4 a four-factor model with two types of language knowledge keeping the same two comprehension factors of Model 3. Out of the four proposed models, this study selects a four-factor model representing L2 listening abilities.

The multifactor listening model of this study supports theoretical models of language abilities. This model speculates that L2 listening consists of attributes of language knowledge and cognitive comprehension. The language attribute has two factors of phonological and illocutionary knowledge which can be extrapolated to organizational and functional components of language knowledge in the CLA model. These language factors are also conceptualized in socio-cognitive perspectives and explained with regard to transferable and context-specific knowledge types. As for the comprehension attribute, this study identifies two other factors based on the degrees of comprehension complexity. Explaining the comprehension factors in connection with strategic use, the study calls for comprehensive conceptualizations of strategic competence in language ability models.

As for the limitations of this study, it is important to note that the selected model of this study is not a comprehensive model of L2 listening and confined by the instrument effect. That is, the listening abilities of this study are inferred from a standardized language test that does not measure interpersonal abilities of language use. Thus, the listening models cannot represent interactional dimensions of L2 listening. Exclusion of interactional factors in the model does not mean insignificance or absence of those attributes in L2 listening, but primarily due to instrument effects. Thus, for future research of listening models, it would be desirable to use various tasks that may enable listeners to manifest their interpersonal and social competence and incorporate them into the modeling.

Finally, statistical results can be powerful and hence, should be interpreted with caution. The use of fit indices may not satisfy an absolute evaluation criterion of models. That is, the overall model fit is part of model evaluations. Although the model fit indices used in this study are highly recommended and functional for the purpose of the current study, in order to examine the sophistication of the selected model, it may be necessary to scrutinize the model parameters including factor loadings (λ), measurement error (δ) covariances, and factor covariance (ϕ) estimates, and other statistical values of the model.

Subsequently, it calls for future investigations to examine relations of observed variables and latent variables, and other error terms.

Biostatement: Minhee Eom is Assistant Professor of English at the University of Texas-Pan American. She teaches undergraduate and graduate courses in Applied Linguistics and ESL educations. Her research interests include second language theories, language testing, research methodologies, and language teacher professionalization.

References

Anderson, J. R. (2004). *Cognitive psychology and its implications* (6th ed.). Macmillan.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.

Brown, G. (1990). *Listening to spoken English* (2nd ed.). London: Longman.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Buck, G. (1995). How to become a good listening teacher. In D. J. Mendelsohn and J. Rubin (Eds.), *A guide for the teaching of second language listening* (pp. 113-131). San Diego: Dominie Press, Inc.

Canale, M. (1983). On some dimensions of language proficiency. In Oller, J. W. Jr. (Ed.) *Issues in language testing research* (pp. 333-342). Rowley, Mass: Newbury House.

Canale, M., & Swain, M. (1980) Theoretical bases of communicative approaches to second language leaching and testing. *Applied Linguistics*, *8*, 67-84.

Carroll, J. B. (1983). Psychometric theory and language testing. In J. W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 80-107). Rowley, Mass.: Newbury House.

Celce-Murcia M, Dörnyei, Z., & Thurrell, S. (1995). A pedagogical framework for communicative competence: A Pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6 (2), 5–35

Chalhoub-Deville, M. (2003). Second language interaction: current perspectives and future trends. *Language Testing*, *20*, 369-383.

Chalhoub-Deville, M., & Deville, C. (2005). A look back at and forward to what language testers measure. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 815-831). Mahwah, NJ: Erlbaum.

Crystal, D. (2004). *The Cambridge encyclopedia of the English language* (2nd ed.). Cambridge: Cambridge University Press.

Educational Testing Service (1995). *TOEFL sample test* (5th ed.). Princeton, NJ: Author.

Eom, M. (2010). Two factors of L2 listening. *The Modern Journal of Applied Linguistics*, *2* (4). 295-305.

Faerch, C., & G. Kasper. (1983). Plans and strategies in foreign language communication In C. Faerch and G. Kasper (Eds.), *Strategies in Interlanguage Communications* (pp. 20-60).London: Longman.

Firth, A., & Wagner, J. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *Modern Language Journal*, *81*(3), 285-300.

Freedle, R., & Kostin, I. (1998). Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing*. *16* (1), 2-32.

Hatcher, L. (1994). *A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling.* SAS Institute Inc. NC: USA.

Hymes, d. H. (1972). On communicative competence. In J. B. Pride and J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin.

Joreskog, K. G., & Sorbom, D. (1993). *Lisrel 8 user's reference guide*. Chicago: Scientific Software International.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49, 294–303.

Kline, R. B. (2005). *Principles and practice of structural equation modeling.* New York: Guilford Press.

Kostin, I. (2004). *Exploring item characteristics that are related to difficulty of TOEFL dialogue items*. TOEFL Research Reports, RR-79. Princeton, NJ: Education Testing Service.

Lantolf, J. (1996). SLA theory building: letting all the flowers bloom! *Language Learning*, *46*(4), 713-749.

Lantolf, J. (2000). Introducing sociocultural theory, in J. Lantolf (Ed.), *Sociocultural theory and second language learning.* Oxford press.

Lantolf, J. P., & Appel, G. (1994). Theoretical framework: an introduction to Vygotskian perspectives on second language research in J. P. Lantolf and G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 1-32). Ablex Publishing Corporation, New Jersey.

Leung, C., & Lewkowicz, J. (2006). Expanding Horizons and Unresolved Conundrums: Language Testing and Assessment. *TESOL Quarterly*, *40* (1), 211-234

McNamara, T. F. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics, 18*, 46-46.

Meyer, B. J., & McConkie, G. W. (1973). What is recalled after hearing a passage? *Journal of Educational Psychology, 65* (1), 109-117.

Mendelsohn, D. J (1994). *Learning to listen: a strategy based approach for the second language learner*. San Diago: Dominie Press.

Nissan, S. DeVincenzi, F., & Tang, L. (1996). *An analysis of factors affecting the difficulty of dialogue items in TOEFL listening comprehension*. TOEFL Research Report, No. 51. Princeton, NJ: Educational Testing Service.

Oller, J. (1983). Evidence for a general language proficiency factor: an expectancy grammar. In J. Oller (Ed.). *Issues in language testing* (pp. 3-10). Rowley, MA: Newbury House.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning Strategies in Second Language Acquisition.* Cambridge: Cambridge University Press.

O'Malley, J. M., Chamot, A. U., & Kupper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, 10 (4), 418-437.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Boston: Heinle.

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, *30* (1), 36-56.

Phakiti, A. (2008). Construct validation of Bachman and Palmer's (1996) strategic competence model over time in EFL reading tests. *Language Testing, 25* (2), 237–272

Richards, J. (1983). Listening comprehension-approach, design, procedure. *TESOL Quarterly*, *17*, 189-220.

Rost, M. (2002). Teaching and Researching Listening. London, UK: Longman.

Rost, M. (2005). L2 listening. In Hinkel, E. (Ed.), *Handbook of research in second language teaching and learning* (pp. 503-28). Mahwah, NJ: Erlbaum Associates.

Skehan, P (1998). *A Cognitive Approach to Language Learning.* Oxford: Oxford University Press.

Vandergrift, L. (2003). Orchestrating Strategy Use: Toward a Model of the Skilled Second Language Listener, *Language Learning*, *53* (3), 463–496

Vandergrift, L. (2005). Relationships among Motivation Orientations, Metacognitive Awareness and Proficiency in L2 Listening. *Applied Linguistics*, *26* (1), 70-89.

Young, R. E. (2000). *Interactional competence: challenges for validity*. Paper presented at the annual meeting of the American Association for Applied Linguistics and the Language Testing Research Colloquium, 11 March 2000, Vancouver, British Columbia, Canada.