Research Note: A Model for Receptive and Expressive Modalities in Adult English Learners' Academic L2 Skills

Author: Hye K. Pae, Ph.D.
University of Cincinnati, U.S.A.
September 2012

1. Introduction

Language skills are what make up the ability to comprehend incoming information and articulate outgoing messages. Linguistic processes involve a wide range of intricate systems, comprising morphology (minimal units of meaning), lexicon (words), phonology (the sound pattern of language), syntax (the sentence pattern of language), semantics (the meaning of language), contexts (connections between prior input and current input), plausibility (connections between input and acceptability), and pragmatics (language usage). This complexity radically increases when it comes to second language (L2) learning, simply because it involves more than one language.

L2 input is a requisite in the development of linguistic competence in L2. In second language acquisition (SLA), a question remains as to how much linguistic input gets retained and how overall L2 skills are L2 output dependent. Receptive skills are developed through retained input (i.e., intake; VanPatten, 2004). Receptive skills comprise listening and reading. L2 proficiency is also manifested though the expressive modality comprising speaking and writing. All these linguistic components are not mutually independent but interconnected.

Recent research has indicated that overall linguistic skills comprise multi-trait or multi-componential subskills (Bachman & Palmer, 1981; Sasaki, 1996; Shin, 2005), as opposed to a unitary component hypothesis which posits that language skills consist of a single canonical ability (Oller, 1979). Despite this consensus on multi-dimensional components involved in L2 performance, it is not known how many subskills are involved in L2 proficiency. At the same time, although receptive and expressive skills have been characterized separately, it is still unclear how the factor structures of receptive and expressive language skills, are related each other. Reves and Levine (1988) indicated that listening and reading are distinct but similar skills sharing commonalties in integrated and holistic comprehension of the message. Further research studies have suggested crossover and overlap between listening and reading abilities beyond their unique trait of skill operation and function (Bae & Bachman, 1998).

An examination of group differences has been one of the popular analysis methods employed in the literature. Research has found different factor structures between low- and high-ability groups (Oltman, Stricker, & Barrows, 1988; Swinton & Powers, 1980). Stricker, Rock, and Lee (2005) investigated the factor structure of a prototype of the Test of English as a Foreign Language (TOEFL) iBT using a multiple-group CFA. Through an item parceling procedure, they tested a correlated two-factor model, utilizing one factor for Speaking and the other, a combination of Reading, Listening, and Writing, for three language groups (Arabic, Chinese, and Spanish). They found equal factor loadings and error variances but each of the three groups showed differences between the two factors. In a similar line, Shin (2005) tested four models for three ability groups (low, intermediate, and high) with a cross-equality imposition, and found that a second-order factor model did not support the hypothesis of increasing or decreasing factor differentiation characterized by an examinee's proficiency. He contended that the language backgrounds of examinees might be the source of the measurement variance.

A number of research studies have shown statistical differences across gender. For example, significant differences of performance by gender were found in computer-based tests (Gallagher, Bridgeman, & Cahalan , 2002), item differential estimates in tests of English proficiency as a foreign language (FL; Ryan & Bachman, 1992), L2 comprehension and vocabulary learning in a video-based computer-assisted language learning program (Lin, 2011), and task performance in a tape-mediated assessment of speaking (Lumley & O'Sullivan, 2005).

Purpose and Research Questions

In spite of a wealth of research studies in L2 language and reading skills, little is known about the factor structure of adult English language learners' (ELLs) academic English skills with respect to the relation between receptive and expressive skills. This study investigated the relations between receptive and expressive academic English skills using structural equation modeling (SEM). A theoretical objective of this study was to test whether adult ELLs' observed data support the converging evidence of a multicomponential model of English skills and whether this model determines the extent to which the measurement and structural equation models fit the subsamples of gender and ability. Two research questions were addressed in this study.

1. To what extent are overall L2 skills defined by receptive skills or expressive skills?
2. How well did the grouping variable (i.e., ability and gender subgroups) meet the assumption of the equal latent variable model?

2. Method

2.1 Participants

The participants were 585 adult ELLs1 from 62 countries. Their mean age was 25 years and ranged from 17 to 59. Females accounted for 54.2% (317 examinees) and males 45.8% (268 examinees).

2.2 Measures

The database used for this study was part of the larger field test administered by the Pearson Test of English Academic (PTE Academic). PTE Academic is a computer based English test of English for Academic purposes The measure assesses different skill domains, including four discrete, independent, domains (listening, speaking, reading, and writing) and five integrated-skill domains (listening/speaking, listening/reading, listening/writing, reading/writing, and reading/speaking). Receptive skills were defined as listening, reading, and listening/reading, whereas expressive skills were defined as speaking, writing, listening/writing, reading/writing, and reading/speaking. Although it may be questionable whether the shared-skill domain can be used as one of the latent variables, a decision rule was made based on the output modality. For example, as seen in the expressive construct, reading/writing constitutes a combined, interactive skill, but reading was instrumental to output a writing product. Therefore, the reading/writing indicator was considered an expressive construct.

---

[1] It is not known whether the participants speak English as an SL or an FL. For the sake of convenience and consistency with the literature, however, L2 is used to refer to FL in this paper.

2.3 Observed-Variable Building and Model Testing

Despite concerns about item parceling (i.e., item parceling can mask the factor structure of the data; West, Finch, & Curran, 1995), the use of item parcels has been a common practice in SEM and has been considered more powerful than item use (Bandalos & Finney, 2001). Item parceling was recommended when items share a unidimensional psychometrical nature (Bandalos & Finney, 2001) and when the composite score is normally distributed, because item parcels conform more closely to the assumptions of theory-based estimation methods (Enders & Bandalos, 1999; Bandalos & Finney, 2001). In an item analysis of the test items, unidimensionality of the PTE Academic items was sustained (see Pae, 2012, for more information). Using the item parceling method, nine observed variables were created and entered into the equation of SEM. Since the items showed congruence within unidimensional domains, no distortions of the factor structure and biased parameter estimate were expected.

In accordance with these theoretical principles four competing models were tested in order to choose the best fit model for the data. The four models included a single-factor model, a correlated two-factor model, an independent two-factor model, and a second-order factor model.

3. Results

The means, standard deviations, correlation coefficients of the variables are shown in Table 1. There was a wide range of variability across the variables, suggesting that the data were free from a range-restriction problem. There were significant, positive correlations between the variables, ranging from $r = .40$ to $r = .77$. The highest correlation was found between listening/speaking and listening/writing variables ($r = .77$), followed by listening/writing and listening/reading skills ($r = .74$). The speaking and reading/speaking indicators showed the lowest correlations ($r = .40$).

Table 2. Correlation Matrix

Variable	1	2	3	4	5	6	7	8	9	Mean	SD
1.Listen	1									7.01	2.53
2.Speak	.41	1								2.13	1.70
3.Read	.62	.43	1							22.43	5.27
4.Write	.52	.45	.54	1						3.69	1.51
5.Listening &Speaking	.62	.62	.56	.60	1					41.33	13.90
6.Listening &Reading	.59	.41	.57	.55	.62	1				9.08	3.80
7.Listening & Writing	.70	.52	.65	.68	.77	.74	1			31.00	11.71
8.Reading & Writing	.63	.41	.68	.57	.61	.61	.73	1		9.47	3.62
9.Reading & Speaking	.43	.40	.43	.44	.59	.49	.50	.44	1	8.62	4.28

Note: SD = Standard Deviation; All correlations are significant at the 0.01 level.

A preliminary check for order condition showed that the model was over-identified, meaning that the degrees of freedom were positive. In the testing of the four models, a correlated two-factor model was chosen as the baseline model for a group comparison. The reasons for this choice were as follow: First, the correlated two-factor model was consistent with both theoretical and empirical bases that linguistic abilities were complex and multifaceted. This is consistent with the findings of previous studies (Bachman, & Palmer, 1981; Sawaki, Stricker, Oranje, 2009; Shin, 2005). Secondly, the model fulfilled the parsimonious principle, because it took the smallest number of error-covariance correlations for the tenable model. In the two-factor model (i.e., receptive and expressive latent constructs), receptive skills were indicated by three observed variables, and the expressive skill factor was indicated by six indicators, as seen in Figure 1.

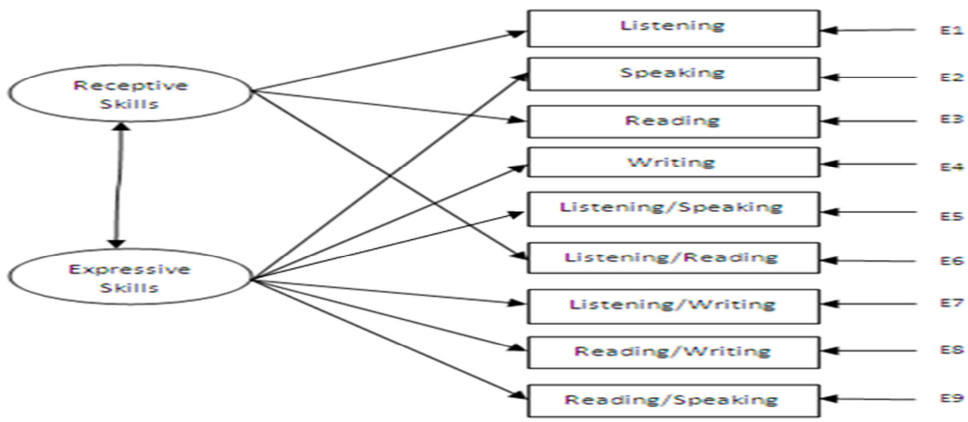

Figure 1. A Correlated Two-Factor Model

The initial model did not fit well [$\chi 2$ (26, N = 585) = 216.95, p = .000, CFI = .97, NFI = .97, RMSEA = .11]. For model modification, the modification indices (MI) and the standardized residual matrix were used. The MI suggested that error covariance be added to observed variables. After a model modification which consisted in adding 7 error-covariance correlations (Listening/Speaking - Speaking; Reading/Writing – Reading; Reading/Speaking - Listening/Speaking; Reading/Speaking - Listening/Writing; Reading – Listening; Reading/Speaking – Speaking; Reading/Writing – Listening), the two-factor model became tenable [$\chi 2$ (19, N = 585) = 30.21, p > .05, CFI = 1.00, RMSEA = .032]. This indicated that the sample covariance matrix S was sufficiently reproduced by this theoretical model. Despite possible criticism of having error terms covary, the addition of the error covariance to the model would be better than adding new paths to the model. This is because an addition of new paths creates different conceptual and theoretical models.

The indicators which were loaded on the receptive factor showed more homogeneous factor loadings in magnitude than those loaded on the expressive construct. When it comes to expressive skills, the range of factor loadings produced by the indicators was wide, and the speaking indicator showed the lowest standardized factor loading, followed by the reading/speaking variable.

3.1 Multiple-Group Model

The whole group was broken down into two male and female subgroups as well as high- and low-performer subsamples. The general procedure was to test for measurement invariance between the unconstrained models for all groups combined, then for a model where certain parameters are constrained to be equal between the groups. The measurement model was used to assess whether or not the factor loadings of the measurement model were invariant across the two gender and ability subgroups. This was accomplished by testing the null hypothesis that the factor loadings were identical across gender and ability subsamples against the alternative hypothesis that the factor loadings were not identical across gender and ability subgroups.

First, the gender subsamples were tested with the two-factor model. The results indicated different parameter estimates for the female and male data when applied to the path model. The female and male data fit the path model as indicated by the non-significant chi-square values ($\chi2= 47.16$, $\chi2 = 19.65$, $ps > .05$, respectively). The global fit statistics indicated a chi-square for the hypothesis of unequal (separate) parameter estimates to be a good fit to the path model ($\chi2= 66.81$, $df = 48$, $p > .05$).

Since it was of interest to test the hypothesis that the two groups shared a common path model, a subsequent SEM was run by setting the parameters specified in the equations to be equal between the two groups. In this analysis, the covariance among the observed variables was free to vary while constraints were placed on the others. The parameter estimates were the same in both groups. The individual chi-square values for each group also summed up to the global chi-square statistic for this common model. The chi-squares for the female and male groups were $\chi2= 43.09$ and $\chi2 = 27.97$, $ps > .05$, respectively, which yield the global chi-square value of 71.06, $df = 61$, $p > .05$. These results indicated that the two sets of data fit the path model based on the hypothesis of similar path coefficients in the baseline path model.

Next a chi-square difference test (a.k.a., a likelihood ratio test) was computed between the two path model analyses. A chi-square difference between the unequal parameter estimates and the equal parameter estimates was calculated using an EXCEL spreadsheet chi-square difference calculator program provided by LISREL 8.80 (Joreskog & Sorborn, 2007). The difference in the chi-square values and associated p-values was obtained using the global chi-square value from the analysis of equal parameter estimates and the global chi-square value from the analysis of unequal parameter estimates. There was no significant chi-square difference between the two model analyses ($\chi2=4.25$, $df =13$). This implies that the female and male data separately fit the path model as well as both datasets fit a common path model.

The group equivalence of models for paths and latent variables for the two ability subgroups (i.e., high- and low-ability groups) were compared as with the gender subgroups. The same model was first applied to the pooled covariance matrices of the two groups in which parameter estimates were averaged across the two groups. Next, the imposed constraints were placed so that parameters could be identical across groups, and then compared to ascertain whether the constrained models with equal parameters fit the data.

The chi-squares of unequal parameter estimates for the low- and high-ability groups were $\chi2= 76.52$ and $\chi2 = 98.23$, $p < .001$, respectively, which yield the global chi-square value of 174.75, $p = 48$, $p < .001$. These results indicated that the two sets of data did not fit the path model based on the hypothesis of similar path coefficients in the baseline path model. With constraints imposed for equal parameter estimates, the low- and high-ability groups' data did not fit the path model as indicated by the significant chi-square values ($\chi2= 91.05$, $\chi2 = 146.86$, $p < .001$, respectively). The global fit statistics indicated a chi-square for the hypothesis of equal parameter estimates to be misfit to the path model ($\chi2= 237.91$, $df = 62$, $p < .001$).

The results of the chi-square difference indicated a difference between the two model analyses ($\chi 2=63.16$, df =14, $p > .05$). The large p value ($p = 3.251$) suggests that there was sufficient evidence that the cross validation of the measurement model for the nine observed variables across the ability subsamples was not supported by the data. Since the ability multigroup CFA produced a misfitting model, no further reports are provided.

4. Discussion

A correlated two-factor model suggested that the two latent constructs, receptive and expressive skills, accounted for adult ELLs' observed test scores. The correlated two-factor model was congruent with the multicomponential model of language skills documented in previous research (Backman et al., 1995; Sawaki, Stricker, & Oranje, 2009; Shin, 2005). Unlike the findings of previous research (Sawaki, Stricker, & Oranje, 2009; Shin, 2005), however, the data did not support a second-order factor model. It seems that the second-order factor has a relatively weak influence on the receptive and expressive skills characterized by the ELLs' performance scores.

The greater variance explained by the receptive skills might have stemmed from the way that L2 speaking is influenced by both linguistic and affective factors. As Song (2008) notes, it is possible that international adult ELLs demonstrate a mismatch between receptive and expressive language skills, typically showing higher written-language proficiency than spoken-language skills. Alternatively, it might have to do with the characteristics of academic language skills. Academic English involves cognitively demanding tasks that require higher-order problem-solving skills and critical reasoning, and uses low-frequency vocabulary. There might be considerable overlaps between academic content and task demands in academic language. Finally, it might have to do with the modality difference. Receptive skills and expressive skills may be represented differently because of different processing routes in which primary language input processing becomes intake (i.e., retained input) to the linguistic competence repository and finally for the efficient mechanism of production

Multiple-group analyses in covariance-based structural equation modeling provide useful information about invariance across subgroups. The chi-square difference test showed that gender subgroups did not differ in the factor structure in relation to the receptive and expressive latent constructs, indicating that the gender-grouping variable did not have an influence on the factor loadings for the observed variables of the data. This compatibility of the data with the model confirms that the subskill factor structure of receptive and expressive linguistic components are similar across gender. This finding is important in the field of adult ELLs' SLA. When it comes to ELLs' academic English skills, the latent constructs of males' and females' achievement seem to be comparable.

With respect to ability groups, there was sufficient evidence that factor structure for the high-proficiency and low-proficiency groups was different. Therefore, a decreasing or increasing differentiation effect as a result of the proficiency level could not be specified, as Shin (2005) noted. This finding suggests that there might be a different latent construct for different ability groups; that is, the factor structure for a highly functioning group is different from that for a poorly performing counterpart. Hence, caution needs to be taken in the interpretation of test scores according to the skill level.

The theoretical implication of their findings has to do with the identification of underpinning factor structures of receptive and expressive L2 skills by adult ELLs across gender and ability subgroups. Since the correlated two-factor model was tenable, the results point to pedagogical implications with respect to the interrelated constructs of input and output modalities. The findings of this study provide evidence for establishing the foundation for effective instructional programs that take both receptive and expressive skills into account .

References

Bachman, L. F., & Palmer, A. S. (1981). The construct validation of the FSI oral interview. Language Testing, 31, 67-86.
Bandalos, D., & Finney, J. S. (2001). Item parceling issues in structural equation modeling. In A. G. Marcoulides & E. R. Schumacker (Eds.), New developments and techniques in structural equation modeling (pp. 269-296). Hillsdale, NJ: Lawrence Erlbaum.
Bae, J., & Bachman, L. F. (1998). A latent variable approach to listening and reading: Testing factorial invariance across two groups of children in the Korean/English two-way immersion program. Language Testing, 15, 380-414.
Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (eds.), New developments and techniques in structural equation modeling (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum.
Enders, K. C., & Bandalos, L. D. (1999). The effects of heterogeneous item distribution on reliability. Applied Measurement in Education, 12, 133-150.
Gallagher, A., Bridgeman, B., & Cahalan, C. (2002). The effect of computer-based tests on racial-ethnic and gender groups. Journal of Educational Measurement, 39, 133-147.
Lin, L-F. (2011). Gender differences in L2 comprehension and vocabulary learning in the video-based CALL program. Journal of Language Teaching and Research, 2, 295-301.
Lumley, T., & O'Sullivan, B. (2005). The effect of test-taker, audience and topic on task performance in tape-mediated assessment of speaking. Language Testing, 22, 415-437.
Oller, J. (1979). Language tests at school: A pragmatic approach. London: Longman.
Oltman, P. K., Stricker, L. J. & Barrows, T. S. (1988) Native language, English proficiency, and the structure of the Test of English as a Foreign Language. TOEFL Research Report Series. Princeton, NJ: Educational Testing Service.
Pae, H. K. (2012). A psychometric measurement model for adult English language learners: Pearson Test of English Academic. Educational Research and Evaluation, 18, 211-229.
Reves, T., & Levine, A. (1988). The FL receptive skills: Same or different? System, 16, 327-336.
Ryan, K., & Bachman, L. F. (1992). Differential item functioning on two tests of EFL proficiency. Language Testing, 9, 12-29.
Sasaki, M. (1996). Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analysis. New York, NY: Peter Lang.
Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. Language Testing, 26, 5-30.
Shin, S-K. (2005). Did they take the same test? Examinee language proficiency and the structure of language tests. Language Testing, 22, 31-57.
Song, M-Y. (2008). Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. Language Testing, 25, 435-464.
Stricker, L. J., Rock, D. A., & Lee, Y-W. (2005). Factor structure of the LanguEdge test across language groups (TOEFL Monograph Series MS-32). Princeton, NJ: Educational Testing Service.
Swinton, S. S., & Powers, D. E. (1980). Factor analysis of the Test of English as a Foreign Language for several language groups (TOEFL Research Report No. 6). Princeton, NJ: Educational Testing Service.
West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, Issues, and applications (pp. 56-75). Thousand Oaks, CA: Sage.
VanPatten, B. (2004). Input and output in establishing form-meaning connections. In B. VanPattern, J. Williams, S. Rott, & M. Overstreet (eds.), Form-meaning connections in second language acquisition (pp. 29-47). Mahwah, NJ: Lawrence Erlbaum.