Linguistic Research 33(Special Edition), 1-27

Relationships among metacognitive skills, listening, and academic reading in English as a foreign language*

Hye K. Pae[a] · **Rose A. Sevcik**[b] · **Daphne Greenberg**[b] · **Sun-A Kim**[c]
(University of Cincinnati[a] · Georgia State University[b] · The Hong Kong Polytechnic University[c])

Pae, Hye K., Rose A. Sevcik, Daphne Greenberg, and Sun-A Kim. 2016. Relationships among metacognitive skills, listening, and academic reading in English as a foreign language. *Linguistic Research* 33(Special Edition), 1-27. This study examined relationships among metacognitive skills, including *inference, summarizing skills, fluency and memory,* listening, and academic reading in English as a foreign language (FL). As a secondary analysis of the field test of the *Pearson Test of English Academic*, a total of 585 nonnative speakers' academic language and reading skills were analyzed using confirmatory factor analysis. One-factor and four-factor models were tested using 12 observed variables and four latent constructs for structural equation modeling to establish a model of reading in English as an FL. The model of reading was tenable for nonnative speakers' skills in English for academic purposes. Consistent with previous findings, *listening* skills were found to be important for reading skills. Other metacognitive skills, such as *inference, summarizing,* and *fluency and memory,* were also robust predictors of efficient academic reading. Of the given variables, the most dominant variable in FL reading for academic purposes was *fluency and memory*. **(University of Cincinnati · Georgia State University · The Hong Kong Polytechnic University)**

Keywords reading, model, nonnative speakers of English, academic English, listening, metacognitive skills

1. Introduction

In the development of a native language (first language; L1), reading acquisition

* This study was supported in part by the Pearson Test of English Academic and the University Research Council at the University of Cincinnati to the first author. The authors would like to thank the anonymous reviewers for their valuable comments on the earlier version of this article.

** Corresponding Author

takes a different trajectory from that of oral language. The former needs to be explicitly taught because it requires a conscious effort to gain the skills (Adams 1990), while the latter comes naturally upon adequate exposure and use from early ages, and is largely effortless. The effortful learning processes of reading have prompted researchers to make attempts to investigate the mechanism underlying L1 reading processes for decades. When it comes to reading in a second language (L2) or a foreign language (FL), it becomes more complicated, simply because an additional language is involved, resulting in L1 facilitation and interference. A number of cross-language research studies have shown convergent evidence of cross-linguistic skill transfer (Gottardo, Yan, Siegel, and Wade-Woolley 2001; Ho and Bryant 1997; McBride-Chang, Bialystok, Chong, and Li 2004; Pae, Sevcik, and Morris 2010). Phonological processing skills and other high-level cognitive processes, such as strategic problem-solving skills, are shown to be language universal (McBride-Chang 2004), but specific skills related to the script and orthography are language specific (Perfetti and Liu 2005). The aim of this study was to build a theoretical model of reading for nonnative speakers (NNSs) of English as FL[1]. Specifically, essential subskills which were identified for proficient reading in L1 and L2 were tested using confirmatory factor analysis (CFA) in order to build an optimal pathway to NNSs' reading in English. This inquiry is crucial because no solid theoretical model of reading in adult NNSs' academic reading as FL has been established and because research shows that theoretical models tend to be valid for both L1 and L2 learners (Verhoeven and Leeuwe 2012). Bernhardt (2005) also called for a reading model because models are "economical" (p. 142), lay foundations to move forward, and facilitate the further development of L2 reading research. Moreover, little research has investigated the within-language relationship in FL, although L1-L2 cross-language transfer has been well documented in the literature (Gottardo, Yan, Siegel, and Wade-Woolley 2001; Ho and Bryant 1997; McBride-Chang, Bialystok, Chong, and Li 2004; Pae *et al*. 2010).

1.1 Theoretical models of reading

Theoretical models of reading point to variables significantly contributing to reading outcomes. A number of reading models, such as the simple view of reading

[1] For the sake of convenience and consistency with the literature, FL and L2 are used interchangeably in this paper unless the reference is related to the learning context.

(Gough and Tunmer 1986), the convergent skills model of reading development (Vellutino, Tunmer, Jaccard, and Chen 2007), the construction-integration model (Kintsch 1994, 1998), and the direct and inferential mediation (DIME) model of reading comprehension (Cromley and Azevedo 2007), have been proposed for English monolingual readers. The simple view of reading posits that reading comprehension (RC) is the product of decoding (D) and listening comprehension (LC), and the relationship is formulated in the equation of RC = D x LC (Gough and Tunmer 1986). In this model, symmetry between the two componential skills is emphasized. The formula indicates that, if either D or LC constitutes a lack of proficiency (i.e., zero), no reading comprehension (i.e., zero) occurs. The complexity and irregularity between spoken words and their written representations in English demand automaticity of decoding because spending too much time in decoding results in no cognitive resources left for reading comprehension (Wolf and Bowers 1999). Along with effortless, quick, and accurate decoding, listening comprehension has been supported as a critical component in other studies for the pathway to successful reading (Neubaus, Roldan, Boulware-Gooden, and Swank 2006). Research has also indicated that listening comprehension is influenced by working memory and cognitive resources (Just and Carpenter 1992). The model of simple view of reading has been expanded to L2 reading comprehension with robust findings for the applicability of this model to L2 (Proctor, Carlo, August, and Snow 2005; Verhoeven and Leeuwe 2012).

Vellutino and his colleagues (2007) have also expanded the simple view model of reading by proposing a convergent-skills model of reading development. A series of reading subskills and reading-related cognitive abilities, such as visual coding, phonological coding, visual analysis, phonological awareness, semantic knowledge, syntactic knowledge, spelling, context-free word identification, language comprehension, and reading comprehension, were tested to identify multivariate components for reading ability. The hypothesized structural model pointed to a convergent-skills model of reading development with phonological skills, semantic and syntactic skills, and visual skills as the preeminent determinants of reading competence. In a study of relationships among reading skills of struggling adult readers, Sabatini and his colleagues (2010) have found that the simple view model of reading is tenable without adding vocabulary and fluency factors to the model. This highlights the importance of listening comprehension, indicating that word reading and listening comprehension alone adequately account for the variance

in reading comprehension.

Kintsch (1988, 1994, 1998) has outlined the architecture of a reading comprehension process in the construction-integration model on the basis of psycholinguistic theories about human memory and text comprehension processes. According to him, a reader creates an understanding of the text in his/her mind while reading, and builds three different mental representations of the text, including (1) a verbatim representation of the text, (2) a semantic representation which conveys the meaning of the text, and (3) a situational representation of the context to which the text refers. The low-level association between contextual information and long-term memory is constructed during reading and is, in turn, utilized to control knowledge activation via a constraint-based integration process. The inferential and propositional representations are derived from the text and are transformed into a network of propositions. If the text is coherent, all nodes of the network are connected to one another such that the meaningful construction of the text becomes efficient.

The componential skills identified in children's reading are developmentally bound. Once readers pass the linguistic threshold of reading, higher-level cognitive strategies play a part in text comprehension. Cromley and Azevedo (2007) tested fewer observed variables than those used in Vellutino *et al*.'s study (2007), including background knowledge, strategies, inference, word reading, vocabulary, and reading comprehension, to propose the DIME model. The DIME model specifies direct and mediated effects of those indicators on comprehension, which explains 66% of the variance in academic reading comprehension. This model has been tested and verified with domain-general text at the high school and undergraduate level. Cromley and Azevedo (2007) have noted a wide application of their model, suggesting that the effects of the given variables on reading performance are qualitatively the same for struggling adolescent readers and for more proficient readers.

1.2 Reading in L2 or FL

A handful of studies examined L2 reading comprehension (Brantmeier 2006; Karimi 2015, 2016; Lee 2014; McNeil 2012; Nassaji 2007) from various angles from cross-language transfer to an L2 schemata theory. McNeil's (2012) study is in line with shared linguistic skills between L1 and L2 (Karimi 2015; Verhoeven and Leeuwe 2012) in that 50% of the variance in L2 reading was explained by L2

knowledge and L1 reading skills. Karimi (2015, 2016) found a dominant role of strategic processing in both L2 single-text and multiple-texts comprehension. Lee (2014) also reported a significant contribution of overall L2 proficiency measured using TOEFL or TOEIC to L2 reading comprehension. Nassaji (2007) employed the schema theory and knowledge-based processes to explain the nature of reading comprehension, but his study was not backed up with empirical evidence.

A number of strategies, such as identifying the main idea, summarizing, drawing inferences, generating questions, creating visual images, and looking for clues, have been identified as effective reading strategies for adults' reading (Hock and Mellard 2005). Of these strategies, summarizing and drawing inferences are considered most important reading comprehension strategies for adult reading outcomes (Hock and Mellard 2005). The importance of these skills becomes more critical in English for academic purposes such that cognitive, linguistic demands are higher than in social language use. Gleaned from previous research in both children's and adults' reading, the skills of listening comprehension, summarizing, memory, inference are considered crucial components in reading.

In addition to the complex process involved in L2 reading, reading for academic purposes imposes an additional complexity to the task. Reading is different from language use in social contexts, and reading for academic purposes is different from that for entertainment purposes. Reading for academic purposes involves a wide range of specific skills of text comprehension, including lexical, syntactic, and discourse processes. Unlike reading for enjoyment, reading for academic purposes requires psycholinguistic strategies, including making predictions and inferences as well as metacognitive strategies, such as previewing, scanning for gist, sequencing, attending to details, and connecting presented information to existing schemata of knowledge. In order to gain an understanding of what they read, readers typically self-monitor text comprehension by means of metacognitive strategies in summarizing, sequencing, comparing/contrasting, inferencing, relating background knowledge to the text, as well as finding the main idea, important facts and details, and supporting evidence.

Reading for academic purposes is inherently challenging given the demand for a higher level of cognitive strategies. When an L2 is involved, its complexity increases to the level above and beyond academic reading. The components of L2 reading ability have been a controversial topic in the literature. One line of research suggests

a global unitary ability in L2 reading, while the other line poses a set of multiple-trait skills. The former, a unitary trait hypothesis, has been supported by Oller (1983) who argues that various components of language skills converge on a general language proficiency factor. The latter, a multiple-trait hypothesis, posits that the construct of reading ability can be divided into a set of multiple componential constructs, possibly in a hierarchical second-order model with lower-level first-order linguistic skills (Alderson 2000; Bachman and Palmer 1981; Bachman, Davidson, and Foulkes 1990; Sawaki 2007; Shin 2005). For example, Sawaki (2007) showed that L2 speaking skills were defined by five subskills and one higher-order ability. The higher-order model has been broadly accepted in the second language acquisition (SLA) community. A number of contributory factors to efficient reading have been empirically identified and validated. The componential approach to reading has noted that reading ability is composed of a set of theoretically distinct and empirically divisible constituents (Hoover and Tumner 1993). However, the scope, the nature, and the extent to which such skills are specifically segmentable still remain controversial.

In spite of a wealth of informative research findings on pathways to reading in English as an L1, few studies have hypothesized and tested a theoretical model of reading for adult NNSs' L2 English for academic purposes. Efforts on L2 reading model were made primarily in children's L2 reading comprehension. For example, Proctor and his colleagues (2005, 2006) expanded theoretical models of reading by adding discrete component skills to produce a model of L2 comprehension. Their hypothesized model was supported by subsequent findings (Proctor *et al.* 2006). Verhoeven and his colleague (2012) also tested the relationship of word decoding and listening comprehension to reading comprehension in both L1 and L2 learners of Dutch. They found that the relationship among word decoding, listening comprehension, and reading comprehension between L1 and L2 was highly comparable, suggesting that the reading model is equally applicable to both L1 and L2 reading. As indicated earlier, there have been studies that examined adults' L2 reading comprehension. However, the difference between previous research and this study involves different variables used in analysis (Karimi 2015, 2016; Lee 2014; McNeil 2012) as well as the use of performance indicators or outcomes rather than self-reported data (i.e., questionnaire; Brantmeier 2006). Hence, a model for adult NNSs' L2 reading for academic purposes is in need for theory building and the

advancement of research in the field.

1.3 The present study

Although the role of isolated word reading, phonological processing, vocabulary, and syntactic knowledge in reading in both L1 and L2 has been well documented in the extant literature (Cromley and Azevedo 2007; Pulido and Hambrick 2008; Vellutino *et al.* 2007), there has been a paucity of research studies on the influence of inferential integration, text analysis, and verbal fluency on FL reading. Of interest is how FL metacognitive skills, operationalized as FL *inference*, *summarizing,* and *fluency and memory*, are associated with FL *reading* skills. A subsequent question would be which construct is most pivotal in predicting reading outcomes. These questions are important for theoretical and practical reasons. The theoretical reason involves a model addressing how adult NNSs' reading in FL is manifested in performing communicative tasks (i.e., listening, speaking, reading, and writing) required in the academic setting. The practical relevance concerns the applicability of the model to pedagogical considerations. The constructs identified as a critical foundation for successful reading in FL will inform practitioners of instructional strategies that foster skillful academic reading outcomes in higher education settings.

Given limited research on the achievement constructs with adult NNSs, this study was motivated to evaluate, using CFA, whether measurement models of latent constructs identified in L1 reading research are viable in adult NNSs' academic reading achievement. Although they are key elements in L2 development, speaking and writing skills are not included in this study because of the focus on reading. In this study, the factor structures of various linguistic skills and the extent to which these observed variables were related to FL academic English reading outcomes of adult NNSs were evaluated using a dataset of the *Pearson Test of English Academic* (*PTE Academic*, 2009), which is a proxy to TOEFL. The significance of this study lies in the fact that speakers of languages other than English take the *PTE Academic* exam for post-secondary or higher education in English-speaking countries and that institutions in which English is a medium of instruction use the test results as a screening measure for admissions. Given the nature of the test (i.e., a proficiency test of academic English), it is worthwhile to build a theoretical model of reading for academic purposes. The following two research questions were examined in this study.

1. What is the underpinning factor structure of the diverse subskills of FL academic English?

Based on the results of previous research, it was hypothesized that the observed variables would positively contribute to FL metacognitive skills as a unitary skill (i.e., a one-factor model) and separable subskills (i.e., a four-factor model).

2. What relationships exist among FL *inference*, *summarizing, fluency and memory, listening* proficiency associated with FL *reading* skills?

Based on the results of previous research that confirmed the importance of listening comprehension (Sabatini *et al.* 2010; Vellutino *et al.* 2007) and its close relationship with reading comprehension (Verhoeven and Leeuwe 2012), it was hypothesized that the three FL latent trait factors (i.e., *inference, summarizing,* and *fluency and memory*) would contribute to FL *listening* skills, and, in turn, to FL *reading* skills.

The rationale for the choice of the variables and the measure was two-fold: (1) The variables under consideration have been consistently identified as fundamental skills in previous research and (2) *PTE Academic* offers an excellent opportunity to build a theoretical model of reading for speakers of English as an FL, as it measured NNSs' English proficiency that is required for academic tasks.

2. Method

2.1 Participants

The participants were adult NNSs of English who were recruited for the Field Test of *PTE Academic*. A total of 585 were selected from the Form 2 database of the first field test. Their mean age was 25 years, and ranged from 17 to 59 years of age. Females accounted for 54.2% (317 examinees) and males 45.8% (268 examinees). About a half of the participants had studied English for more than 10 years (53%) and lived in English-speaking countries (57%). The participants' countries of origin were 62 in total. India had the largest portion of the participant pool, followed by China, Israel, Taiwan, France, and Korea. Although their countries of origin had a wide range, the participants shared commonalities in terms of their

interests (i.e., admittance into academic programs at higher education institutions where English is an instructional medium). Because the analyses were based on the test-taker's academic language and reading skills in English, the homogeneity in the skill domain might precede the variability of their L1s and multilingual status. Previous research conducted using *PTE Academic* data showed no significant differences found between two groups by a 10-year demarcation of English study and between two groups of the participants who had lived in English-speaking countries and those have not (Pae 2012). Therefore, it can be deduced that the sample homogeneity is established in the data.

2.2 Procedure

The participants' test scores were drawn from the larger Field Test database of *PTE Academic*, which was administrated by Pearson worldwide in 2007. The participants were individually tested at the international testing locations provided by Pearson (Pearson 2009). Given that this study was a secondary data analysis, test scores were selected for data analysis from the participant pool available from *PTE Academic* (see below for more information about the *PTE Academic* test).

2.3 Measures

PTE Academic is a high-stakes test and a proxy exam to the TOEFL, which is typically used for a measure of readiness for learning at the college level. It is an international, computer-based, English assessment battery focused on academic English skills in the areas of listening, speaking, reading, and writing. It takes three hours to complete the test battery. Test-takers of *PTE Academic* are individuals who plan to study at English-medium higher education institutions in English-speaking countries. The content of test items includes real-life, authentic audio or video excerpts from academic lectures or textbooks and activities taken place in academic settings. Because of the nature of a timed test, no basal and ceiling rules were established in the test. Form 2 of the Field Test was utilized for this study. Concerning psychometric properties, split-half reliability coefficients for the field test range from .76 to .94 in the domains of reading, listening, writing, and speaking (Pearson 2009). Another construct validity testing revealed that the measure assessed what it purported to measure for the sample, showing Kuder-Richardson 20 (KR 20)

reliability coefficients ranging from .96 to .99 (Pae 2012).

The indicators utilized in this study were chosen as important components of reading ability due to their theoretical relevance based on the aforementioned reading models. These observed variables were defined by five latent constructs: *inference, summarizing, fluency and memory, listening,* and *reading*. Although phonological processing and word identification skills have been identified as precursors to efficient reading in the models, these variables were not included in this study mainly because (1) the participants had already passed the threshold of basic English reading given that all of them finished high school or above and (2) the instrument purported to gauge the test-taker's English skills using academic English materials which required the higher level of language skills than phonological processing and word identification skills.

Inference. This construct consisted of three skills, including *reordering sentences, filling in blanks of reading materials,* and *selecting missing words*, because the skill to make inferences comprises the ability to guess the meaning of unknown words from the text and the context. The task of reordering sentences asked the participant to reconstruct randomly scrambled sentences by moving the sentences and placing them in a coherent order. The stimuli for filling-in-the-blank task presented a gapped real-life reading text, and asked for choosing the word or phrase which best supplied each gap in the text. The task of selecting missing words required the test taker to listen to a real-life, authentic audio excerpt from an academic lecture or conversation and to complete the gapped written text by typing four to seven words missing from the text (Pearson 2009).

Summarizing Skills. This construct was operationalized using tasks of *highlighting correct summary, summarizing written text,* and *summarizing spoken text*, because summarizing requires both local and global comprehension to grasp the gist of the text. The first task asked the test-taker to select the correct paragraph out of three to five paragraphs after viewing a real-life audio or video of an academic lecture or speech. The second task required the participant to read the text and summarize it using one sentence of up to 30 words. The last task involved writing a summary of what the speaker had said in a real-life, authentic audio/video excerpt from an academic lecture (Pearson 2009).

Fluency and Memory. This construct was composed of *reading aloud, dictation,* and *sentence repetition*. The reading aloud task assessed fluency, while the dictation and sentence repetition tasks measured the capacity of working memory. The reading

aloud task asked the participant to read a short text of 40-60 words. Dictation asked the test-taker to listen to a short sentence of 8-11 words, to transcribe what was spoken, and to type the exact sentence in the space provided on the computer screen. Sentence repetition task asked the participant to listen to a short scripted recording and to repeat the sentence exactly as heard (Pearson 2009).

Listening. The listening task presented an audio or video recording about academic subjects and asked the test-taker to answer different types of questions (Pearson 2009). Three different listening question types were included under the construct of *listening* because more than one indicator per latent construct is typically recommended in structural equation modeling (SEM).

Reading. The reading stimuli included academic lectures, academic subjects, and textbook contents (Pearson 2009). Test-takers were asked to respond to two different types of questions, which served as indicators.

2.4 Data analysis

CFA using LISREL 8.80 (Joreskog and Sorbom 2007) was performed to test the hypothesized model. CFA was selected as a statistical technique because it partials out the differences in measurement errors across the observed variables and has the ability to model error terms as a covariance structure analysis, making SEM advantageous over regression modeling. The sample size was large enough to perform SEM. To begin with, the data were checked for missing data, multicollinearity, outlier, and normality assumptions, based on Kline's (1998) suggestions for data preparation and data screening for SEM. There were no missing data because no responses to the question items were coded as incorrect responses, given that the assessment tool was a timed test and that no basal and ceiling rules were applied.

For hypothesis testing, two hypotheses were tested using CFA models. The results were interpreted based on the empirical appropriateness of a model, which was evaluated with respect to its adequacy to explain the data by appropriate statistical means. Analyses were based on covariance matrices with the maximum likelihood method.

To evaluate statistically significant theoretical models, several fit criteria were used for the statistical significance and relevant meaning of the theoretical model. Although the non-statistical significance of the chi-squared likelihood test has been broadly used to appraise a goodness-of-fit in SEM, the chi-squared statistics are sample-size sensitive

and tend to be inflated by large sample sizes and models with many variables (Schumacker and Lomax 2004). In addition to the chi-squared statistics, three criteria were used for the evaluation of theoretical models: (1) the nonsignificance of the $\chi 2$ statistics, the $\chi 2/df$ ratio (3.0 or below, as suggested by Kline, 1998), the comparative fit index (CFI; .95 or above), the goodness of fit index (GFI; .95 or above), the root mean-square error of approximation (RMSEA; .05 or below), and expected cross-validation index (ECVI), (2) the statistical significance of individual parameter estimates for the paths in the model, and (3) the magnitude and the direction of the parameter estimates (Schumacker and Lomax 2004).

3. Results

3.1 Descriptive statistics

The means and standard deviations of the observed variables, along with the indices of skewness and kurtosis, are shown in Table 1. The variables were approximately normally distributed with skewness values ranging from -1.326 to 0.077 and kurtosis values ranging from -1.080 to 1.855. Since there were no standard scores available and the data covered a wide range of values, square-root-transformed scores were also obtained in order to place the scores on the linear, same scale, continuum and to reduce the range to a more manageable range. The means of the square-root-transformed values ranged from 1.00 to 5.95. The scores showed a wide range of variability, and a problem of restricted ranges was not an issue in the data. Two different analyses were performed using square-root-transformed units and raw scores to estimate the factor loadings and test the statistical significance of the correlation coefficients among the constructs under consideration. Since the results were not significantly different across the two analyses, the results using the raw scores were reported in this paper. The raw scores rather than the square-root-transformed units were used for two reasons: (1) a lack of variability in the transformed scores and (2) the lower magnitude of correlations in the square-root-transformed units than the raw scores.

Table 1. Means and standard deviations of the observed variables

Observed Variables	Raw Scores					Square-Root Transformed Scores	
	Range	Mean	SD	Skewness	Kurtosis	Mean	SD
Reorder Paragraphs	0-12	9.07	1.86	-1.047	1.855	2.00	.35
R Filling Gaps	0-13	6.88	2.30	-.257	-.124	2.58	.50
LR Filling Gaps	0-13	9.59	3.24	-1.053	.404	3.02	.66
LR Highlights	0-3	1.74	.94	-.252	-.836	1.22	.50
RW Summarization	0-4	1.39	1.03	.077	-1.080	1.00	.62
LW Summarization	0-3	1.49	.85	-.648	-.638	1.09	.55
Read Aloud	0-15	8.62	4.28	-.380	-.708	2.76	1.01
Dictation	0-36	19.92	8.64	-.193	-.860	4.31	1.15
Sentence Repetition	0-57	36.84	11.98	-.528	.304	5.95	1.20
Listening 1	0-3	2.45	.76	-1.326	1.295	1.53	.34
Listening 2	0-6	2.76	1.55	.077	-.611	1.54	.61
Listening 3	0-3	1.81	.95	-.272	-.910	1.25	.49
Reading 1	0-3	1.85	.82	-.333	-.268	1.30	.41
Reading 2	0-8	4.63	2.15	-.399	-.909	2.06	.61

Note: SD = standard deviation; R = Reading; LW = Listening/Writing; LR = Listening/Reading; RW = Reading/Writing

Bivariate correlation coefficients were computed. All the correlations were significant at the 0.01 level, ranging from .27 to .76. The correlation coefficient between sentence repetition and dictation skills was the highest (r = .76), followed by those between dictation and filling gaps in listening/writing items and between dictation and summarization skills in listening/writing items. The reading outcome variable showed medium strengths of correlations with the other observed variables.

3.2 Model testing: Research question 1

A two-step approach was utilized. The first step was to test the measurement model using CFA to evaluate whether the observed variables served as adequate indicators of the latent constructs. The second step involved the structural model to examine the predictive relationship between the latent constructs and the reading outcome variable.

One-factor and four-factor models were tested to determine whether a reasonably

good fit of the data to the model exists[2]. The first model (one-factor model) hypothesized that one unitary latent construct (i.e., FL global skills) defined all the indicators. The theoretical relevance to this model was a single-component hypothesis which assumed that FL skills were derived from a single canonical entry (Oller 1983). This unitary trait hypothesis indicates that various components of language skills are different manifestations of a single, general, language ability. The four-factor model was also tested, given the multi-componential language-skill hypothesis which has been supported by a number of research studies (Bachman and Palmer 1981, 1982; Bachman, Davidson, and Foulkes 1990; Harley, Cummins, Swain, and Allen 1990; Sawaki 2007; Shin 2005). It was hypothesized that the multiple factor models satisfactorily accounted for the covariances in the data. The four-factor model included *inference*, *summarizing*, *fluency and memory*, and *listening* proficiency. The *listening* construct was set up as a separate trait on the basis of previous research that identified it as a dominant factor in children and adult reading (Gough and Tunmer 1986; Neuhaus *et al.* 2006; Sabatini *et al.* 2010; Vellutino *et al.* 2007).

At the stages of model estimation and model testing, many model solution problems emerged, including matrices that were not positive definite, negative error variances, and poor general fit. Each misfit model was inspected to identify theoretically justifiable relationships for model improvement. The modification indices were reviewed for each model to determine whether adding variables to other factors, deleting variables from the model, or adding error covariances to the model would be appropriate. The first model (i.e., one-factor model) was originally misfitting, but became tenable through model modification by adding eight error-covariance correlations (i.e., Sentence Repetition - Read Aloud; Sentence Repetition − Dictation; Listening Summary - Listening Gap; Dictation − Reorder; Sentence Repetition - Listening Gap; Sentence Repetition − Reorder; Listen3 - Listening Summary; Listen3 - Listening Gap) among the observed variables [$\chi 2$ (46, N = 585) = 66.70, p = .027, CFI = 1.00, RMSEA = .027]. Figure 1 displays the one-factor model. Although there

---

[2] Two-factor and three-factor models were also tested. The two-factor model included constructs of *metacognitive* skills and *listening* proficiency. The three-factor model comprised constructs of *metacognitive* skills, *verbal fluency,* and *listening skills*. The two-factor model was also plausible through the model modification procedure by adding eight error-covariance correlations among the indicators. The three-factor was acceptable by having six error covariances correlated. Since the results were similar to those of the four-factor model, however, the results of two- and three-factor models are not reported in this paper.

is criticism on having error terms covary, the addition of the error covariance to the model did not deviate from the theoretical relevance. From a theoretical point of view, having error terms covary is better than adding a new path to the model, because adding a new path may generate a different theoretical model (Schumacker 2010). Since of interest was to test a pre-formulated theoretical model, double loadings to variables were not allowed in model specification. The result of the first model was consistent with that of previous research (Oller 1983).

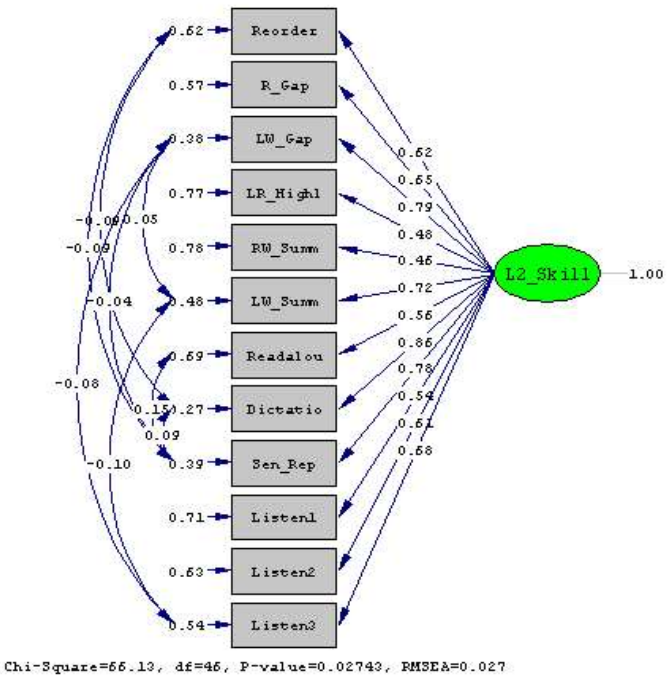

Figure 1. A one-factor model

Note: R = Reading; LW = Listening/Writing; LR = Listening/Reading; RW = Reading/Writing; Readalou = Read Aloud; Dictatio = Dictation; Sen_Rep = Sentence Repetition; Listen1 = Listening 1

With the same procedure taken in the one-factor model, the four-factor model was acceptable by having six error covariances correlated (i.e., Listening Summary - Reading Gap; Dictation - Listening Gap; Dictation - Read aloud; Sentence Repetition

- Read Aloud; Listen3 – Reorder; Listen3 - Listening Summary). Figure 2 shows the four-factor model. The chi-squared values suggested reasonably good fits of the data to the hypothesized models. The GFI index suggested that 98% of the variance-covariance in matrix *S* was reproduced by the models.

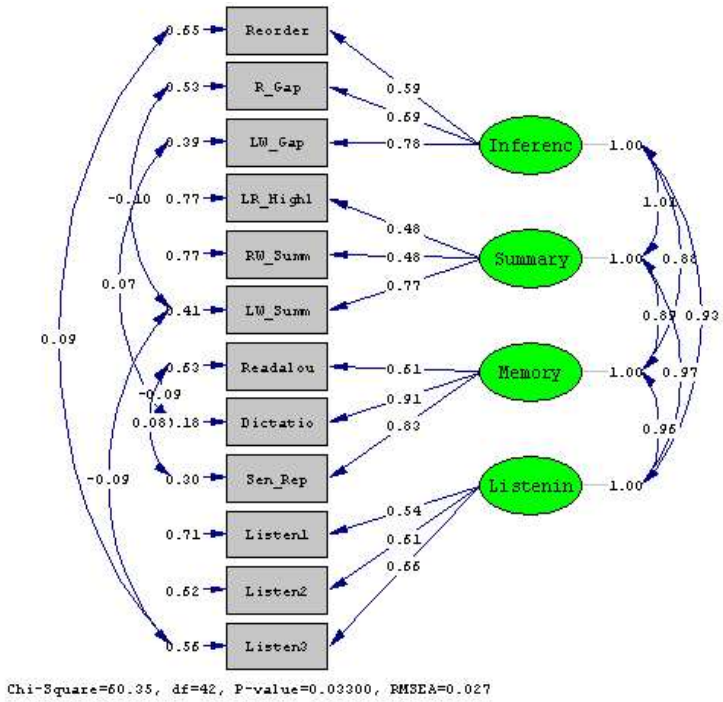

Figure 2. A four-factor model

Note: R = Reading; LW = Listening/Writing; LR = Listening/Reading; RW = Reading/Writing; Readalou = Read Aloud; Dictatio = Dictation; Sen_Rep = Sentence Repetition; Listen, Listenin = Listening

As indicated earlier, the first model required eight error covariances to be correlated in the models, as indicated earlier, while the second models had six error covariances correlated to form good-fit models. In consideration of model parsimony (i.e., the smaller number of error covariances added to the model) and the consistency with the theoretical relevance of previous research, the four-factor model was favored over the one-factor model. The four latent constructs, including

inference, summarizing, fluency and memory, and *listening* skills, defined the observed variables well in the four-factor model. The model indicated that the observed variables were adequate indicators of the latent factors and supported the theoretical model. The fit indices for the four-factor model were within the acceptable range, with a CFI near zero residuals [$\chi 2$ (42, N = 585) = 60.35, p = .03, $\chi 2/df$ = 1.437, CFI = 1.00, RMSEA = .027]. All of the standardized factor loadings from the latent constructs to the indicators were significant. In the four-factor model, the *dictation* indicator score accounted for 82% of the variance in *verbal fluency*, followed by *sentence repetition* explaining 70% of the variance in *verbal fluency*. *Highlighting* task in listening/reading stimulus showed the lowest factor loadings (.48) and the lowest R-squared value (22%; i.e., only 22% of the variance was explained by *highlighting* skills). The intercorrelations among the factors in the model were substantial to high, ranging from .83 to .98.

3.3 Final model: Research question 2

Given the strong theoretical support for *listening* proficiency in the reading model (Gough and Tunmer 1986), despite the comparative low factor loadings and variances explained by the indicator variables in the listening subtests, a revised model was tested. The revised model included three latent constructs (*inference, summarizing,* and *fluency and memory*) explained by the nine indicators, and, in turn, the three constructs defined *listening* skills. Reading ability was regressed on the three exogeneous variables (*inference, summarzing,* and *fluency and memory*) and one endogenous variable (*listening*).

Structural invariance is related to the equivalence of relationship among latent variables. In testing the final model, model specification and convergence problems, such as negative degrees of freedom, non-positive definite matrix, or the Haywood case, were observed. For solutions, error covariances were correlated based on the suggestion of modification indices and, as seen in Figure 3, the variance of the several factors were fixed to 1 for identification purposes and for the purpose of avoiding some serious parameter conditions encountered during optimization in SEM. Figure 3 shows the final model including the standardized path coefficients that indicate the relative importance of relations within this specified model. Reading abilities were defined by four latent factors of *inference, summarizing, fluency and*

memory, and *listening* skills. The fit indices for the hybrid model indicated goodness of fit based on the ratio of the chi-squared value and *df*, CFI, GFI, and RMSEA [$\chi 2$ (62, N = 585) = 117.22, p = .000, $\chi 2/df$ = 1.89, CFI = .99, GFI = .97, RMSEA = .04], indicating that the observed variables were adequately measuring constructs as specified in the model.

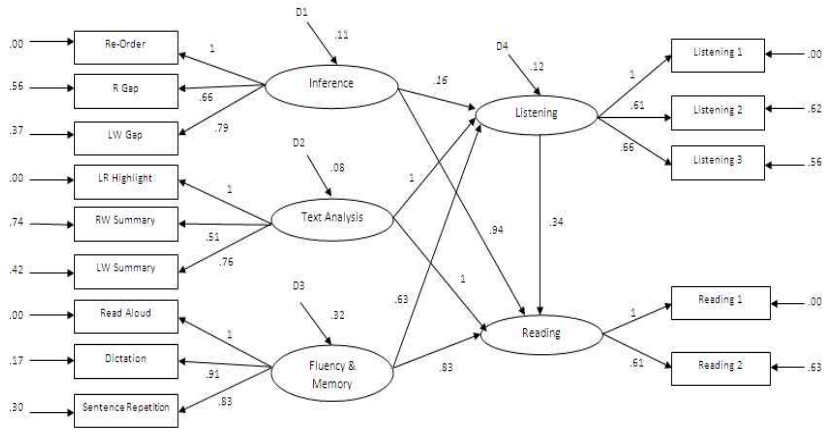

Figure 3. Final model

Note: The factor loading from *Inference* to *Listening* was not significant. The rest of the loadings were significant at the 0.05 level. D = Disturbance

4. Discussion

In reading, many factors, such as pattern recognition, lexical access, concept activation, syntactic analysis, propositional encoding, and inferential integration are simultaneously involved in its cognitive processing. FL reading is a more complex cognitive process than L1 reading, because it involves multiple componential skills from word recognition to meaning construction of the text as well as L1 interference. Various L1 reading models served as a foundation of the proposed model testing for FL English reading. This study was designed to examine the factor structure of FL *inference*, *summarizing*, *fluency and memory* skills, and *listening* proficiency in relation to FL *reading* competence. Of interest was an identification of the

underpinning factor structure and a pathway to L2 academic English skills by adult NNSs. The relative significance of FL *inference, summarizing, fluency and memory,* and *listening* proficiency as predictors of FL reading was examined. The hypothesized model was supported, but asymmetry was found in the comparative contributions made by the subskills and FL reading competence. In essence, the applicability and sustainability of predictive subskills necessary for successful L1 reading were upheld for NNSs' FL English reading.

4.1 Toward a model of academic reading in FL

The relationships of the observed variables and the latent constructs provided a statistically significant and theoretically feasible explanation for FL reading performance of the participants in this study. The results suggested that each of the latent variables accounted for a significant amount of variance in the observed variables. On the basis of the unitary-skill hypothesis (Oller 1983) and the multiple-subskill hypothesis (Bachman and Palmer 1981, 1982; Bachman, Davidson, and Foulkes 1990; Harley, Cummins, Swain, and Allen 1990; Sawaki 2007; Shin 2005), the two models (i.e., one-factor and four-factor models) were tested. With the inclusion of error-term correlations to the model, the two models showed goodness-of-fit. The factor structure underlying the mechanism of FL reading supports one canonical linguistic construct as well as a series of multiple traits involved in FL reading. One explanation for these possible models is the interconnectiveness of linguistic skills. Given that the test items of *PTE Academic* were constructed based on academic contents, skills required to process academic materials may be largely governed by one canonical metacognitive skill set. At the same time, different performance indicators can define different latent traits according to the nature of error covariance correlations involved in the model. Although the present study provided evidence of an interface among the indicators and latent constructs, causality of the relationships is difficult to be determined.

Of the two measurement models, the four-factor model, which was the decomposed, distinguishable, skills model, was chosen mainly because of the theoretical reason to further test a structure model. Model testing identified statistically significant influences of the four separate trait factors on FL academic reading. The paths from the constructs to the indicators of *re-ordering, highlighting,*

and *reading aloud* were constrained to 1 in order to optimize the model. The other variables showed significant factor loadings on the indicators allocated. The *inference* construct exhibited the only nonsignificant factor loading to the *listening* factor, while the other two factors (*summarizing* and *fluency and memory*) showed direct effects on the *listening* skills. The *listening* construct demonstrated both direct and indirect effects on FL academic reading, which is consistent with the notion of the simple view of reading (Gough and Tunmer 1986; Sabatini *et al.* 2010; Verhoeven and Leeuwe 2012). The *inference*, *summarizing,* and *fluency and memory* factors showed direct relationship to the FL *academic reading* outcome. These findings are in line with the results of previous research studies (Cromley and Azevedo 2007; Sabatini *et al.* 2010; Vellutino *et al.* 2007).

A constellation of findings spanning children's L1 English reading to adults' FL English reading has converged in support of FL listening comprehension as a precursor to FL academic reading. The results of this study shows support for the findings of previous studies. In addition to the importance of listening skills to reading achievement, other metacognitive skills, such as *inference, summarizing,* and *fluency and memory,* are also robust predictors of efficient academic reading. The findings of this study confirm FL within-language interactional relationships of multicomponential skills.

4.2 Theoretical implications

From a theoretical perspective, one important finding from this study was that *dictation* subskill was a consistently robust indicator to the latent construct. The modification indices suggested that *dictation* was defined by the *listening* construct, and that the *sentence repetition* indicator was defined by *inference, summarizing,* and *listening* factors. Since this study attempted to test an *a priori* established theoretical model, as indicated earlier, double loadings were not allowed in the paths. A possibility of double loadings indicates possible overlaps between the factors, suggesting multiple relations of the *dictation* and *sentence repetition* indicators to other factors beyond the *a priori* path. Although double loadings were not permitted in model specification, the modification indices suggested that the *listening* construct defined *dictation* and *sentence repetition* indicators. This is consistent with the notion of the simple view of reading, in which listening comprehension is a crucial

component of proficient reading.

Performance on *dictation* and *sentence repetition* is related to working memory span. Both tasks assess how many words the test-taker can accurately write or repeat upon hearing them only once by holding the incoming information for a certain period of time. The ability to retain and retrieve long and linguistically complex sentences requires efficient working memory capacities and appropriate memory span. If a test-taker dictates or repeats short sentences verbatim but omits words in longer sentences, it suggests a possibility of a reduced memory span. If a test-taker makes errors at the end of sentences (a primacy effect) or errors at the beginning of sentences (a recency effect), it suggests problems with working memory. In addition to memory span or working memory problems, inattention or lack of interest at the time can also affect initial encoding of information and the ability to recall what is heard, as indicated by errors on shorter sentences and verbatim responses on longer ones. Although it does not offer specifics about the primacy and recency effects and inattention, this study highlights significant effects of *dictation* and *sentence repetition* subskills on academic English reading. This is also consistent with the results of previous research with children. Further research is warranted to corroborate this finding. It should be noted that the model is not prescriptive of pathways to FL academic reading but descriptive of how NNSs' FL academic linguistic skills are associated with latent constructs.

L2 or FL learning encompasses various aspects of perception, comprehension, storage, and retrieval of information under consideration. These different aspects of learning activities may not necessarily share all their sub-processes. For example, comprehension sometimes requires summarization, simplification, and integration of materials to be learned with prior knowledge or the schemata, while retention and retrieval of information, as required in the task of dictation and sentence repetition, have a close relation with the memory span and mental lexicon storage. The extant body of literature has provided a wide spectrum of theoretical models ranging from decontexualized word reading to highly cognitive processing of reading comprehension. The path model identified in this study suggests that FL metacognitive skills, such as *inference, summarizing,* and *fluency and memory*, are salient predictors of FL academic English reading.

4.3 Methodological implications

Since this study attempted to identify the underpinning factor structure and a plausible path model toward FL academic reading in English, based on the test scores on different skill domains, the results revealed the relationships among outcome measures. Still unknown is the factor structure of task characteristics. A subsequent study can build upon the results of this study. One possible approach may be to include the factor loadings or average factor score for each item analysis to build a task characteristic model (Carr 2006). The factor structure of task characteristics can offer insights into how NNSs approach different tasks as well as strengths and weaknesses of their problem-solving strategies.

This study used NNSs' FL data only. As a result, the discussions of this study do not address cross-linguistic interdependent relationships. If the extent of overlap between L1 and FL performance is identified using both L1 and FL data, it can promote an understanding of inter-language associations. The construct map of academic oral and written skills in L1 and FL can broaden the knowledge base of cross-language relationships, including facilitating or inhibitory L1 effects on FL performance. Since the results of this study outline the factor structure of FL skills, the model tested in this study can be a basis for cross-language construct mapping.

The participants had heterogeneous L1 background and FL learning contexts in which the degree to which they were exposed to English was variable. Based on a consistent pattern of achievement and unidimensionality in their performance scores (Pae 2012), the homogeneity of the sample was assumed. The results of this study invite a cluster analysis to examine whether scores in the same L1 groups are more similar to one another than scores in other L1 groups.

4.4 Pedagogical implications

The results of this study can be integrated into lesson plans for adult NNSs. Given that evidence-based pedagogy is valued in classrooms, the provision of the significant latent variables predicted by observed test scores will inform teachers and practitioners of which construct to highlight in their classroom teaching in order to maximize instructional effectiveness for NNSs. The latent traits of linguistic skills found to be influential in skillful reading can help teachers and practitioners define the scope and breadth of reading instruction for academic purposes.

4.5 Future directions

Previous studies have shown that vocabulary size is an essential factor affecting reading performance in L1 and L2/FL (Vellutino *et al.* 2007). Vellutino *et al*. (2007) found that word semantic knowledge showed factor loadings of .70 for grades 2 and 3 and .86 for adolescent students on language comprehension. The role of L1 and L2 vocabulary in L2 reading also has been emphasized in the literature, but this study did not include a vocabulary variable despite its importance in reading. An inclusion of L2 vocabulary in the path model of L2 reading performance will allow for a more comprehensive reading model in L2. Further research is needed to address this issue.

A comparison of the factor structure between L1 and L2 will provide insights into cross-language differences. This study did not include the participants' L1 due to the unavailability of the mother database. In need is a study that addresses a contrast between the two languages by setting the parameters in L1 and then transferring them to L2.

In spite of numerous discussions on NNSs' survival English and English for academic purposes, there has been no solid model of survival English and English for academic purposes tested in the literature. To our knowledge, this study is the first which examined the factor structure and proposed an FL academic reading model for FL readers. Subsequent research studies are needed to corroborate the findings of this study. A comparison of the L2/FL models between survival English and academic reading also will shed light onto the understanding of basic communicative English skills and cognitively demanding academic English skills.

References

Adams, Marilyn J. 1990. *Beginning to read: Thinking and learning about print*. Cambridge: The MIT Press.

Alderson, J. Charles. 2000. *Assessing reading*. Cambridge: Cambridge University Press.

Bachman, Lyle F. and Adrian Palmer. 1981. The construct validation of the FSI oral interview. *Language Testing* 31: 67-86.

Bachman, Lyle F. and Adrian Palmer. 1982. The construct validation of some components of communicative proficiency. *TESOL Quarterly* 16(4): 449-465.

Bachman, Lyle F., Fred Davidson, and John Foulkes. 1990. A comparison of the abilities measured by the Cambridge and Educational Testing Service EFL Test Batteries. *Issues in Applied Linguistics* 1: 30-55.

Bernhardt, Elizabeth. 2005. Progress and procrastination in second language reading. *Annual Review of Applied Linguistics* 25: 133-150.

Brantmeier, Cindy. 2006. Toward a multicomponent model of interest and L2 reading: Sources of interest, perceived situational interest, and comprehension. *Reading in a Foreign Language* 18(2): 89-115.

Carr, Nathan T. 2006. The factor structure of test task characteristics and examinee performance. *Language Testing* 23(3): 269-289.

Cromley, Jennifer and Roger Azevedo. 2007. Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology* 99: 311-325.

Gottardo, Alexandra, Bernice Yan, Linda Siegel, and Lesly Wade-Woolley. 2001. Factors related to English reading performance in children with Chinese as a first language: More evidence of cross-language transfer of phonological processing. *Journal of Educational Psychology* 93(3): 530-542.

Gough, Philip and William Tunmer. 1986. Decoding, reading, and reading disability. *Remedial and Special Education* 7: 6-10.

Harley, Birgit, Patrick Allen, Jim Cummins, and Merrill Swain (eds.). 1990. *The development of second language proficiency*. Cambridge: Cambridge University Press.

Ho, Connie Suk-Han and Peter Bryant. 1997. Learning to read Chinese beyond the logographic phase. *Reading Research Quarterly* 32: 276-289.

Hock, Mick and Daryl Mellard. 2005. Reading comprehension strategies for adult literacy outcomes. *Journal of Adolescent and Adult Literacy* 49: 192-200.

Hoover, Wesley A. and William Tunmer. 1993. The components of reading. In George Brian Thompson William E. Tunmer, and Tom Nicholson (eds.), *Reading acquisition processes*. Clevedon, PA: Multilingual Matters.

Joreskog, Karl G. and Dag Sorbom. 2007. *LISREL 8.80*. Hillsdale, NJ: Scientific Software International.

Just, Marcel A. and Patricia Carpenter. 1992. A capacity theory of comprehension: individual differences in working memory. *Psychological Review* 99(1): 122-140.

Karimi, Mohammad N. 2015. L2 multiple-documents comprehension: Exploring the contributions of L1 reading ability and strategic processing. *System* 52: 14-25.

Karimi, Mohammad N. 2016. Prior topical knowledge and L2 proficiency as determinants of strategic processing in English for Academic Purposes multi-texts comprehension. *Innovation in Language Learning and Teaching:* 1-12.

Kintsch, Walter. 1988. The use of knowledge in discourse processing: A construction-integration model. *Psychological Review* 95: 163-182.

Kintsch, Walter. 1994. The psychology of discourse processing. In Morton Ann Gernsbacher (ed.), *Handbook of psycholinguistics,* 721-739. San Diego, CA: Academic Press.

Kintsch, Walter. 1998. *Comprehension: A paradigm of cognition.* New York: Cambridge University Press.

Kline, Rex B. 1998. *Principles and practice of structural equation modeling*. New York, NY: Guilford.

Lee, Yong-Geun. 2014. Levels of representation in L2 reading comprehension: Effects of L2 proficiency, prior knowledge, and L1 reading skills. *Foreign Languages Education* 21(2): 1-25.

McBride-Chang, Catherine, Ellen Bialystok, Karen Chong, and Yanping Li. 2004. Levels of phonological awareness in three cultures. *Journal of Experimental Child Psychology* 89: 93-111.

McNeil, Levi. 2012. Extending the compensatory model of second language reading. *System* 40(1): 64-76.

Nassaji, Hossein. 2007. Schema theory and knowledge based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning* 57(Supplement s1): 79-113.

Neuhaus, Graham F., Luis Roldan, Regina Boulware-Gooden, and Paul Swank. 2006. Parsimonious reading models: Identifying teachable subskills. *Reading Psychology* 27: 37-58.

Oller, John W. Jr. 1983. Evidence for a general language proficiency factor: An expectancy grammar. In John Oller (ed.), *Issues in language testing research*, 3-10. Rowley, MA: Newbury House.

Pae, Hye K. 2012. A psychometric measurement model for adult English language learners: Pearson Test of English Academic. *Educational Research and Evaluation* 18: 211-229.

Pae, Hye K., Rose Sevcik, and Robin Morris. 2010. Cross-language correlates in phonological process and naming speed: Evidence from deep and shallow orthographies. *The Journal of Research in Reading* 33(4): 335-436.

Pearson Test of English Academic. 2009. Retrieved August 1, 2012. [http://www.pearsonpte.com/PTEACADEMIC/Pages/home.aspx]

Perfetti, Charles A. and Ying Liu. 2005. Orthography to phonology and meaning: comparisons across and within writing systems. *Reading and Writing: An Interdisciplinary Journal* 18: 193-210.

Proctor, C. Patrick, Maria Carlo, Diane August, and Catherine Snow. 2005. Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology* 97(2): 246-256.

Proctor, C. Patrick, Diane August, Maria Carlo, and Catherine Snow. 2006. The intriguing role of Spanish language vocabulary knowledge in predicting English reading

comprehension. *Journal of Educational Psychology* 98(1): 159-169.

Pulido, Diana and David Hambrick. 2008. The *virtuous* circle: Modeling individual differences in L2 reading and vocabulary development. *Reading in a Foreign Language* 20(2): 164-190.

Sabatini, John P., Yasuyo Sawaki, Jane Shore, and Hollis Scaborough. 2010. Relationships among reading skills of adults with low literacy. *Journal of Learning Disabilities* 43: 122-138.

Sawaki, Yasuyo. 2007. Construct validation of analytic rating scales in a speaking assessment: Reporting a score profile and a composite. *Language Testing* 24: 355-390.

Schumacker, Randall E. 2010. Personal written correspondence.

Schumacker, Randall E. and Richard Lomax. 2004. *A beginner's guide to structural equation modeling* (2nd edition). Mahwah, NJ: Lawrence Erlbaum.

Shin, Sang-Keun. 2005. Did they take the same test? Examinee language proficiency and the structure of language tests. *Language Testing* 22(1): 31-57.

Vellutino, Frank R., William Tunmer, James Jaccard, and RuSan Chen, R. 2007. Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading* 11(1): 3-32.

Verhoeven, Ludo and Jan Leeuwe. 2012. The simple view of second language reading throughout the primary grades. *Reading and Writing* 25: 1805-1818.

Wolf, Maryanne and Patricia Bowers. 1999. The double-deficit hypothesis of the developmental dyslexia. *Journal of Educational Psychology* 91: 415-438.

Hye K. Pae
School of Education
University of Cincinnati
615R Teachers College, Cincinnati, OH 45221, USA
E-mail: hye.pae@uc.edu

Rose A. Sevcik
Department of Psychology
Georgia State University
736 Urban Life Building 140 Decatur Street, Atlanta, GA 30303, USA
E-mail: rsevcik@gsu.edu

Daphne Greenberg
Department of Educational Psychology, Special Education, and Communication Disorders
Georgia State University
PO Box 3979, Atlanta, GA 30302-3979, USA
E-mail: dgreenberg@gsu.edu

Sun-A Kim
Department of Chinese and Bilingual Studies
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
E-mail: ctskim@polyu.edu.hk

Received: 2016. 08. 01.
Revised: 2016. 09. 12.
Accepted: 2016. 09. 12.