



Modeling Relationships Among Young Learners' Self-assessment, Learning Attitude, and Language Test Performance

Antony John Kunnan

University of Macau, Macau

Christine Linyu Liao

University of Macau, Macau

Introduction

In many countries around the world, young learners from as early as six or seven years of age start learning English as a second language (or ESL). But not all such learners are successful language learners in terms of their language proficiency. In fact, in a few years, by the time the young learners are 10 or 11, there is a noticeable difference in ESL proficiency levels. This is the case even when the learners are in the same classroom, in the same school, where the instructor and the instruction provided is by the same teacher and the teaching-learning materials too are the same. The question that can be asked is why this is the case. Are there factors that contribute to successful language learning other than instruction?

This question has been addressed by numerous researchers for a number of decades in second language acquisition (SLA) and assessment. Gardner (1985, 1988) and Clément and Kruidenier (1985) started investigating this question by using a socio-educational modelling of second language learning focusing on individual differences in second language learning in terms of language aptitude, attitude and motivation, attrition, formal and informal learning contexts, cognitive style, intelligence and situational anxiety. While Gardner's model and research studies have enriched our understanding of factors that influence successful SLA, there have been challenges to the general model (Oller, 1982), to the measures of language proficiency (Bachman, 1988), and to modelling techniques (Au, 1988; Bachman & Skehan, 1998; Skehan, 1991, 1998).

Brief Review of the Literature

As mentioned above, researchers have long been interested in exploring the relationships between learner characteristics and language learning achievement. The development and popularity of the structural equation modelling (SEM) approach in the 1990s enabled researchers to investigate learner characteristics and independent latent constructs (along with relevant observed variables) and language proficiency as dependent latent constructs (along with observed variables). For example, Kunnan (1994, 1995) investigated the relationship among test takers' home country instruction, English-speaking country

exposure, monitoring, and their language test performance using the SEM approach. Purpura (1997), Phakiti (2003), and Zhang, Goh, and Kunnan (2014) explored relationships between test takers' strategy use and reading test performance using the same method. More recently, in a path-analytic study, Huang, Hung, and Hong (2016) modelled the relationship among test takers' topical knowledge, anxiety, and their performance on an integrated speaking test.

The relationship of self-assessment, and test scores have gained much attention from researchers due to its double role as a learning and an assessment tool (Boud, 1995). Falchikov and Boud (1989) conducted a meta-analysis investigating the relationship between undergraduates' self-awarded scores and teacher-awarded scores. The correlation and degree of agreement varied greatly across disciplines and studies, which indicates that even undergraduates may not be able to accurately assess their own learning. In the field of language testing, Ross (1998) also found considerable variation in the accuracy of self-assessment through a meta-analysis. Such a result indicates that it might be a challenge for young learners to accurately evaluate their language proficiency. Although many studies point to the value of self-assessment, investigations of self-assessment among young learners in primary school is rare as such students are considered to be too young to evaluate their performance accurately and precisely.

Due to the dearth of relevant research in these areas, little is known about the impact of young learners' learning attitude on their language test performance and the accuracy of their self-assessment. Therefore, this study modelled the relationships among young learners' learning attitude, self-assessment, and language test performance using the SEM approach. Specifically, two research questions were investigated:

1. What is the relationship between young learners' learning attitude and language test performance?
2. What is the relationship between young learners' self-assessment and language test performance?

Methods

Participants

The study participants were 398 young learners from eight schools in four Chinese speaking cities (Hong Kong, Taipei, Taichung, and Kaohsiung). In each city, 99 or 100 students from two grades (Grade 5 and Grade 6) in one or two public/government schools participated in the study. There were altogether 5 and Grade 6) in one or two public/government schools participated in the study. There were altogether 200 Grade 5 students and 198 Grade 6 students. Table 1 presents demographic information about the participants.

TABLE 1
Demographic Information about the Participants

Site	Grade 5	Grade 6	Total
Hong Kong	49	50	99
Kaohsiung	50	50	100
Taichung	50	50	100
Taipei	51	48	99
Total	200	198	398

Instruments

The study used two instruments: (1) a bilingual survey and (2) the General English Proficiency Test for Kids (GEPT-Kids).

A bilingual survey

A bilingual survey in English and Chinese (with both traditional and simplified scripts) developed by the Language Training and Testing Center (LTTC), Taipei, was used to collect students' background information, self-assessment of their abilities, and their attitudes to English language learning. The survey had three sections. *Section 1* had 10 can-do statements requiring students to evaluate their listening (Items 1-5) and reading skills (Item 6-10) on a 1-4 Likert-scale like continuum, with 1 representing 'all of them' and 4 representing 'only a few of them.' *Section 2* had 11 questions about student family background and English learning experience. *Section 3* had 5 statements requiring students to indicate their attitudes towards English and different language skills on a 1-4 continuum, with 1 representing 'strongly agree' and 4 representing 'strongly disagree.' In this study, Section 2 items were not used as they were not relevant to the research questions. Thus, the 15 items in Sections 1 and 3 were renamed from Q1-Q15 (see Appendix). Therefore, Q1-Q5 were about students' self-assessment on listening, Q6-Q10 were about students' self-assessment on reading, and Q11-15 were about students' attitudes towards English language learning.

GEPT-Kids

GEPT-Kids is a pencil-and-paper based English language test designed for young learners by LTTC, Taipei. The test used in this study included two sections: listening and reading. The listening section included four tasks and the reading section included three tasks. The test tasks were in the selected response format, such as multiple-choice, cloze, matching, and judging. The LTTC website has more details: <https://www.geptkids.org.tw/ENHome>.

Data Collection

After trialing the survey and the test in a primary school in Zhuhai, China, with 97 students, the instruments were revised and prepared for the study. Data collection was conducted by administering the survey and the test to the study participants. In each city, one Grade 5 and one Grade 6 class section participated in data collection. The participants were administered the test first which was followed by the survey. Together, the data collection took about 60 minutes. The test and survey were scored by LTTC staff. The test responses were scored dichotomously and coded as 0 or 1 for unacceptable and acceptable answers respectively and item scores for the same task were parceled together for data analysis. The survey responses were coded in reverse: 1 → 4; 2 → 3, 3 → 2, 4 → 1. Therefore, in the analysis of this study, 1 represents 'strongly disagree' or 'only a few of them' and 4 represents 'strongly agree' or 'all of them'.

Data Analysis

Of the data collected from 398 participants, 35 cases included missing values, representing 8.8% of the total cases. As the missing values were found to be missing at random, they were removed from the dataset. Thus, there were 363 cases with valid and complete data. The data was then used to conduct descriptive statistical analyses, and CFA (Confirmatory Factor Analysis) and SEM (Structural Equation Modelling) analyses to explore the relationships between test taker characteristics (i.e., self-assessment and learning attitude) and test performance in GEPT-Kids (listening and reading sections). Prior to CFA and SEM, descriptive analysis and reliability analysis were conducted to check the distribution and the reliability of the data. These analyses were conducted using SPSS24, and CFA and SEM were conducted with AMOS24.

Results

Descriptive Statistics

Table 2 summarizes descriptive statistics for the data. From Table 1, we can see that the means for survey items were quite high (above 3.0), indicating that test takers evaluated their learning attitude and language abilities very positively. The means for test tasks was also very high, indicating that the test was easy for the test takers. The skewness and kurtosis values were beyond -2 and 2, indicating that the data were not normally distributed. One solution was the drop the items and tasks that were above the acceptable kurtosis values (Q2, Q4, Q8, LT1, and LT3) but this procedure was not used as this would have resulted in a smaller set of survey items and test tasks. Another solution is to keep these items and tasks but use an empirical technique termed “bootstrapping” which is a nonparametric approach to estimate standard errors by repeated resampling from the sample data. This technique was used to perform CFA and SEM as suggested by Byrne (2016). The Cronbach’s alpha for the survey responses and test scores ranged from 0.81 to 0.93, indicating high internal reliability of the data.

TABLE 2

Descriptive Statistics for Survey and GEPT-Kids (n = 363)

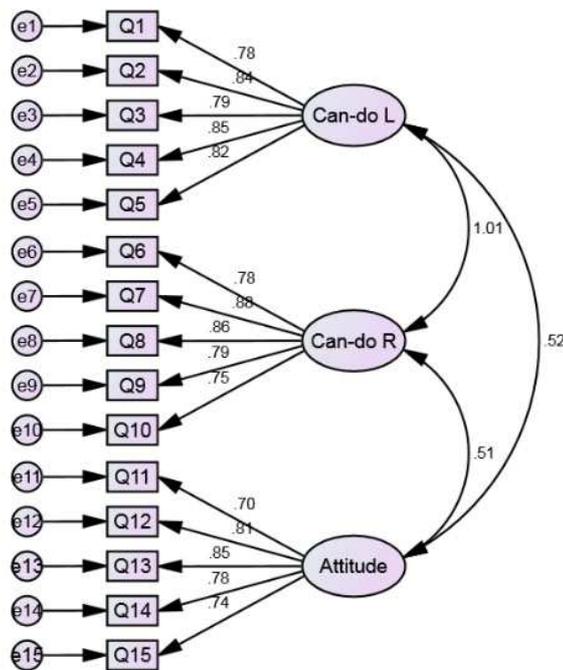
	Min	Max	Mean	SD	Skewness	Kurtosis	Cronbach’s alpha
Q1	1	4	3.58	0.71	-1.66	2.07	
Q2	1	4	3.86	0.44	-3.62	14.67	
Q3	1	4	3.59	0.73	-1.83	2.78	
Q4	1	4	3.84	0.49	-3.66	14.47	
Q5	1	4	3.75	0.60	-2.62	6.65	
Q6	1	4	3.61	0.73	-2.00	3.53	
Q7	1	4	3.79	0.55	-2.77	7.22	
Q8	1	4	3.83	0.49	-3.27	11.62	
Q9	1	4	3.44	0.82	-1.38	1.12	
Q10	1	4	3.56	0.73	-1.77	2.77	
Q11	1	4	3.03	0.87	-0.75	0.02	
Q12	1	4	3.35	0.72	-1.09	1.38	
Q13	1	4	3.36	0.74	-1.18	1.40	
Q14	1	4	3.52	0.68	-1.41	1.87	
Q15	1	4	3.12	0.85	-0.88	0.37	
All items							0.93
LT1	1	5	4.75	0.62	-3.13	11.47	
LT2	1	6	5.61	0.77	-2.45	7.23	
LT3	0	7	6.79	0.79	-5.90	41.40	
LT4	0	7	5.83	1.46	-1.38	1.47	
Listening tasks							0.81
RT1	9	20	18.36	2.20	-2.02	4.73	
RT2	1	5	4.48	0.93	-1.81	2.37	
RT3	0	5	4.01	1.37	-1.34	0.83	
Reading tasks							0.86

Confirmatory Factor Analysis (CFA)

As we were interested in modeling the relationships among young learners’ learning attitude, self-assessment, and language test performance, we proceeded directly with CFA and SEM. According to the survey design, there were three sets of items related to test taker characteristics: self-assessment of listening abilities (Can-do L items Q1-5), self-assessment on reading abilities (Can-do R items Q6-10), and attitudes towards English language learning (items Q11-15). The GEPT-Kids used to examine the study participants’ language test performance had two test sections: Listening (tasks LT1- LT4) and Reading (tasks RT1-RT3). Therefore, a three-factor model for test taker characteristics and a two-factor model for test performance was examined through CFA, which laid the foundation for SEM analysis.

In CFA and later SEM, Bootstrapping ML (Maximum Likelihood) approach and multiple indices were used to estimate model fit. These indices included TLI (Tucker-Lewis index), CFI (Comparative Fit Index), RMSEA (Root Mean Square Error of Approximation), and SRMR (Standardized Root Mean Square). The following standards represent acceptable model fit: TLI and CFI over 0.90, RMSEA less than 0.08 (Heubeck & Neil, 2000), and SRMR lower than 0.08 (Hu & Bentler, 1999). Chi-square (χ^2) was reported as common practice, but it was not used as a model fit index by itself as it is too sensitive to sample size (Marsh, Balla, & McDonald, 1988).

Based on the criteria listed above, the two CFA models did not show to be clearly acceptable models. However, in Figure 1, the regression weights and correlation coefficients of the two models were significant at the 0.001 level, and the factor loadings of all the variables were quite high. Factor loadings of all the survey items were above 0.70, indicating that the factors were well represented by those items. However, as shown in Figure 1, the three-factor model based on the survey showed that self-assessment on listening and self-assessment on reading was very strongly correlated ($r = 1.01$). According to Figure 2, factor loadings of all the GEPT-Kids test tasks ranged from 0.49 to 0.80, indicating that the two underlying factors (listening and reading abilities) were reasonably well represented by the test tasks. The correlation between listening and reading was also very strong ($r = 0.91$).



Notes. Can-do L: Listening related can-do statements, representing self-assessment on listening
 Can-do R: Reading related can-do statements, representing self-assessment on reading

Figure 1. A three-factor CFA model of survey items.

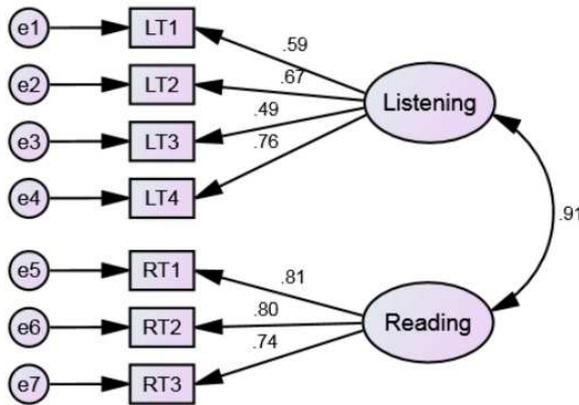
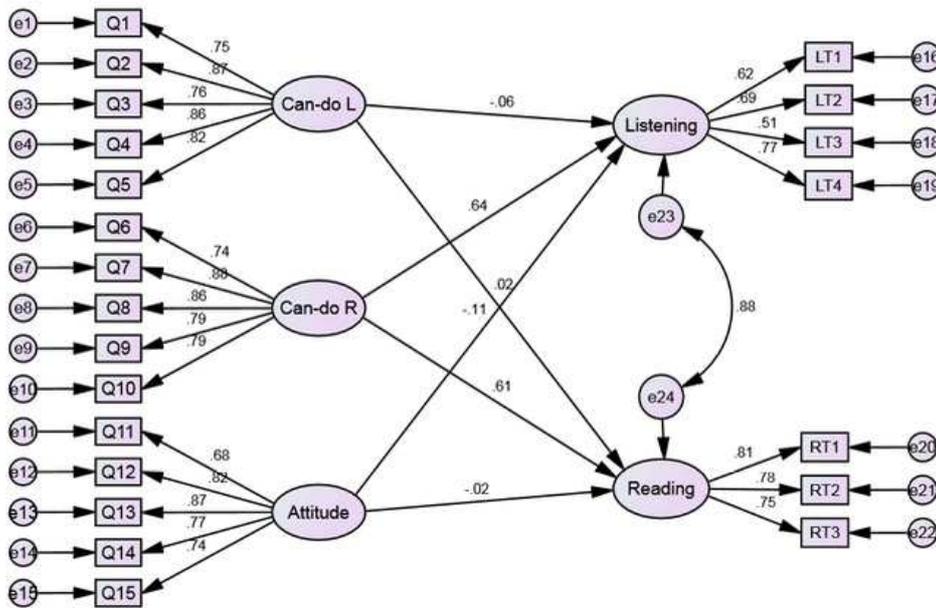


Figure 2. A two-factor CFA model of GEPT-Kids.

Structural Equation Modelling (SEM)

Figure 3 shows the SEM model based on the three test taker characteristics related factors (independent factors) and the two test-performance related factors (dependent factors). Since the two dependent factors (i.e., Listening and Reading) were highly correlated as shown in Figure 2, their error terms were covaried in SEM. The overall model fit of this full model was not acceptable based on relevant indices (TLI: 0.86, CFI: 0.91, RMSEA: 0.14; SRMR: 0.06) and some insignificant regression weights. In addition, from Figure 2, we can see that self-assessment on listening (Can-do L) and learning attitude played almost no role in influencing Listening (-0.06) and Reading (0.02) scores. Therefore, these two factors were removed from further analysis.

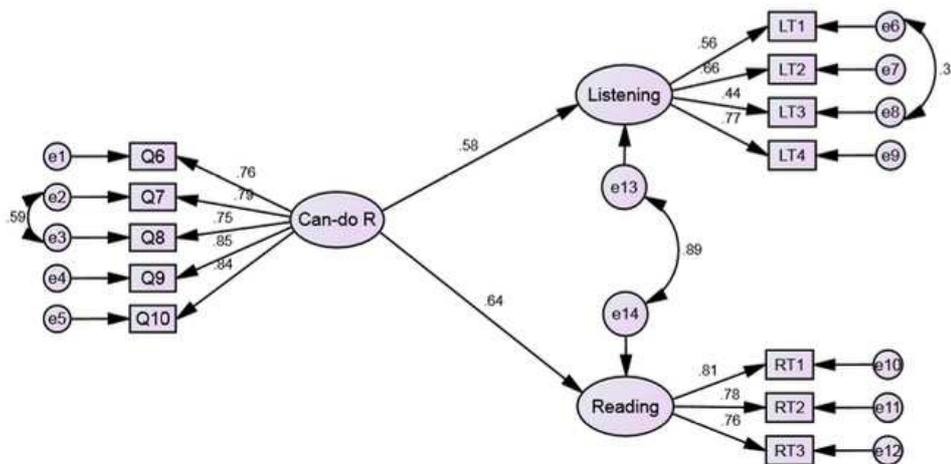


Notes. E23 and E24 were covaried because listening and reading abilities were highly correlated ($r = 0.91$, see CFA in the last section).

LT: Listening Task variables; RT: Reading Task variables

Figure 3. Model 1 with survey and test tasks.

After removing the two factors (Can-do L and Attitudes) that did not have significant influence, a model with a single independent factor (Can-do R) and two dependent factors (Listening and Reading) was analyzed. In this model, all the estimates were significant at the 0.001 level and the standard regression weights between self-assessment and listening/reading ability were moderately high. However, the model did not show an acceptable overall model fit (TLI: 0.83, CFI: 0.86; RMSEA: 0.13, SRMR: 0.06). Therefore, model modifications were made by covarying error terms with high covariance (i.e., e2 & e3, e6 & e8). After adding these modifications, the model (Model 2a) showed a much better model fit (TLI: 0.93, CFI: 0.95, RMSEA: 0.09; SRMR: 0.04). The standardized regression weights in this model show a moderate positive relationship between the participants' self-assessment (Can-do R) and Listening and Reading scores.



Notes. LT: Listening Task variables; RT: Reading Task variables

Figure 4. Model 2 and 2a with Can-do Reading survey items and test tasks.

Table 3 summarizes the SEM factors and fit indices of the above-mentioned models. From Table 3, it is clear that among the three models, the modified Model 2a shows the best model fit. While RMSEA was slightly high, all other indices for the modified Model 2 were acceptable.

TABLE 3
Summary of Models with Fit Indices

Model	Independent Factor(s)	Dependent Factors	Model fit indices				Evaluation
			TLI	CFI	RMSEA	SRMR	
Model 1 (Figure 3)	Can-do L Can-do R Attitudes	Listening Reading	0.86	0.91	0.14	0.06	Not acceptable
Model 2 (Figure 4)	Can-do R	Listening Reading	0.83	0.86	0.13	0.06	Not acceptable
Modified Model 2a (Figure 4)	Can-do R	Listening Reading	0.93	0.95	0.09	0.04	Acceptable

Discussion

According to the results of this study, first, learning attitude had little impact on students' English language test performance although the young learners generally had very good learning attitude or high learning motivation, which could be seen from their high agreement with the statements such as they like

English and they consider language skills important. Perhaps, the high similarity among young learners' learning attitude may have contributed to the lack of influence in their English language test performance.

Next, although young learners are often seen as unable to accurately evaluate their own learning (Butler & Lee, 2010) and some studies showed low correlation between self-assessment and teacher assessment or test scores even among mature learners (cf. Falchikov & Boud, 1989; Ross, 1998), this study found a quite high positive relationship between young learners' self-assessment and their English language test performance. This shows that primary school students, young as they are able to fairly accurately evaluate their own listening and reading abilities. Such accuracy of self-assessment might be due to two reasons. Firstly, as can be seen from the average scores of the test tasks, the young learners' English language proficiency was high. It is understandable that learners with higher proficiency can evaluate themselves more accurately because they know what they know or what they do not know. This may not be the case with lower-level proficiency young learners. Secondly, what the young learners assessed were their receptive skills (i.e., listening and reading). According to Ross's (1998) meta-analysis, self-assessment of receptive skills tends to be more accurate than self-assessment of productive skills among adult learners. As this study only explored young learners' self-assessment of receptive skills, it is not known how accurately they could assess their productive language skills.

Conclusion

Through CFA and SEM analyses, this study modelled relationship among young self-assessment, learners' learning attitude, and language test performance. It was found that young learners' learning attitude had little impact on their language test performance and that there was a quite high positive relationship between young learners' self-assessment and their English language test scores. Since young learners can quite accurately evaluate their own language proficiency, it is feasible and advisable to use self-assessment as an alternate formative assessment tool for the purpose of regulating and monitoring young learners' language learning in primary school.

The main limitation of this study includes the potential bias in the selection of study participants. Since all the participants in this study were high-level proficiency learners from developed cities, the result of this study may not be able to be generalizable to a wider population such as lower-level proficiency students or those from less-developed areas. In addition, language proficiency investigated in this study was limited to listening and reading abilities with the selected response format. To overcome these limitations, it is suggested that future studies investigate a wider population and include productive skills as well.

Acknowledgment

This paper reports on research using examination data provided by the Language Training and Testing Center, Taiwan. This study was supported/funded by the Language Training and Testing Center, Taiwan.

The Authors

Antony John Kunnan is a professor in the Department of English at the University of Macau.

Department of English
Faculty of Arts and Humanities
University of Macau
Email: akunan@umac.com

Christine Linyu Liao is a PhD student in the Department of English at the University of Macau.

Department of English
Faculty of Arts and Humanities
University of Macau
Email: yb77701@umac.mo

References

- Au, S. Y. (1988). A critical appraisal of Gardner's social psychological theory of second language (L2) learning. *Language Learning, 38*, 75-99.
- Bachman, L. F. (1988). Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition, 10*, 149-164.
- Bachman, L. F., & Skehan, P. (Ed.) (1998). *Interfaces between second language acquisition and language research*. Cambridge, U.K.: Cambridge University Press.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*, 5-31.
- Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). London: Routledge.
- Clément, R., & Krudener, B. G. (1985). Aptitude, attitude and motivation in second language proficiency: A test of Clément's model. *Journal of language and Social Psychology, 4*, 21-37.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*, 395-430.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Sage.
- Gardner, R. C. (1988). The socio educational model of second language learning: Assumptions, findings, and issues. *Language learning, 38*, 101-126.
- GEPT-Kids. [n.d.] LTTC, Taipei. Retrieved from <https://www.geptkids.org.tw/ENHome>
- Heubeck, B., & Neill, J. (2000). Confirmatory factor analysis and reliability of the mental health inventory for Australian adolescents. *Psychological Reports, 87*, 431-440.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.
- Huang, H. T. D., Hung, S. T. A., & Hong, H. T. V. (2016). Test-taker characteristics and integrated speaking test performance: A path-analytic study. *Language Assessment Quarterly, 13*, 283-301.
- Kunnan, A. J. (1995). *Test taker characteristics and test performance: A structural modeling approach*. Cambridge, UK: Cambridge University Press.
- Kunnan, A. J. (1994). Modelling relationships among some test-taker characteristics and performance on EFL tests: An approach to construct validation. *Language Testing, 11*, 225-250.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological bulletin, 103*, 391-410.
- Oller, J. W. (1982). Gardner on affect: A reply to Gardner. *Language Learning, 32*, 183-189.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing, 20*, 26-56.
- Purpura, J. E. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning, 47*, 289-294.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experimental factors. *Language Testing, 15*, 1-19.
- Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition, 15*, 275-298.

- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, U.K.: Oxford University Press.
- Zhang, L., Goh, C. M. & Kunnan, A. J. (2014) Analysis of test takers' metacognitive and cognitive strategy use and EFL reading test performance: A multi-sample SEM approach. *Language Assessment Quarterly*, 11, 76-102