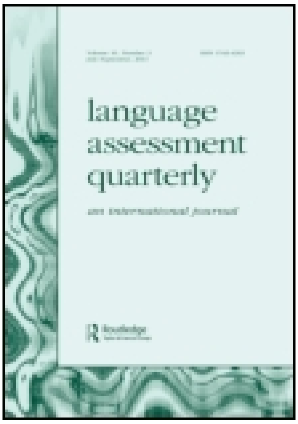

Language Assessment Quarterly
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hlaq20

Analysis of Test Takers' Metacognitive and Cognitive Strategy Use and EFL Reading Test Performance:A Multi-Sample SEM Approach

Limei Zhang [a] , Christine C. M. Goh [a] & Antony John Kunnan [a]
[a] National Institute of Education, Nanyang Technological Institute , Singapore

Published online: 28 Feb 2014.

Language Assessment Quarterly, 11: 76–102, 2014

ISSN: 1543-4303 print / 1543-4311 online
DOI: 10.1080/15434303.2013.853770

Routledge
Taylor & Francis Group

Analysis of Test Takers' Metacognitive and Cognitive Strategy Use and EFL Reading Test Performance: A Multi-Sample SEM Approach

Limei Zhang, Christine C. M. Goh, and Antony John Kunnan
National Institute of Education, Nanyang Technological Institute, Singapore

This study investigates the relationships between test takers' metacognitive and cognitive strategy use through a questionnaire and their test performance on an English as a Foreign Language reading test. A total of 593 Chinese college test takers responded to a 38-item metacognitive and cognitive strategy questionnaire and a 50-item reading test. The data were randomly split into two samples ($N = 296$ and $N = 297$). Based on relevant literature, three models (i.e., unitary, higher order, and correlated) of strategy use and test performance were hypothesized and tested to identify the baseline model. Further, cross-validation analyses were conducted. The results supported the invariance of factor loadings, measurement error variances, structural regression coefficients, and factor variances for the unitary model. It was found that college test takers' strategy use affected their lexico-grammatical reading ability significantly. Findings from this study provide empirical and validating evidence for Bachman and Palmer's (2010) model of strategic competence.

Researchers in language testing have shown interest in the identification and characterization of individual characteristics that influence performance on language tests (Kunnan, 1995; Phakiti, 2008; Purpura, 1997). Most recently, Bachman and Palmer (2010) argued that test takers' metacognitive strategies determine how language ability is actualized in language use. In addition, cognitive strategies, as one of test takers' peripheral attributes, can also affect test performance when language users employ them to "execute plans" (p. 43) in test contexts.

Similarly, studies in reading comprehension have also attached increasing emphasis to the role of strategy use in reading comprehension (Mokhtari & Sheorey, 2002; Pressley & Afflerbach, 1995). Researchers argued that in the meaning-constructing process of reading comprehension, metacognition, with its toolbox of strategies, plays the role of a problem solver, repairing comprehension failure and maximizing comprehension (Pearson, 2009). The general consensus is that strategic awareness and monitoring of comprehension distinguish skilled readers from unskilled ones (Grabe, 2009; Grabe & Stoller, 2002; Paris & Jacobs, 1984). Further research in this respect has addressed the relationship between readers' strategy use and their reading performance (e.g., Carrell, 1989; Phakiti, 2003, 2008).

Correspondence should be sent to Limei Zhang, English Language and Literature Academic Group, National Institute of Education, 1 Nanyang Walk, Singapore 637616. E-mail: limei.zhang@stdmail.nie.edu.sg

In addition, researchers have long been interested in investigating learners' strategy use on language tests (Bachman & Palmer, 2010; Cohen & Upton, 2006; Phakiti, 2008; Purpura, 1999, 2013). In spite of an array of research conducted in this area, no conclusive evidence has been produced regarding the complicated relationships between test takers' strategy use and their test performance. In addition, no empirical studies have been conducted to examine how test takers' metacognitive and cognitive strategies are related in specific use situations, though Bachman and Palmer (2010) incorporated cognitive strategies into their strategic competence model. The present study, therefore, was designed to address the research gap regarding how metacognitive and cognitive strategy use is related to each other in test contexts. Furthermore, it investigates how test takers' strategy use affects their reading test performance using a multi-sample structural equation modeling (SEM) approach. Findings from this study are expected to have theoretical, methodological, and pedagogical implications for language testing and second/foreign language acquisition.

LITERATURE REVIEW

Metacognitive and Cognitive Strategy Use in Relation to Language Use

According to Flavell (1979), metacognition is "knowledge and cognition about cognitive phenomena" (p. 906). In other words, metacognition refers to language learners' ability to think about how they engage in information processing and how they analyze, evaluate, and manage the way they do it (Vandergrift & Goh, 2012). In the context of language use, metacognition or metacognitive awareness is often used as a general term to refer to language learners' awareness and consciousness in adopting appropriate strategic behaviours and activities to solve problems in their cognitive activities related to language use (e.g., Paris, Wasik, & Turner, 1991; Vandergrift, Goh, Mareschal, & Tafaghodtari, 2006).

As noted by Flavell (1979), metacognition plays an important role in many cognitive activities regarding language use (see Goh, 1998, 2008; Vandergrift et al., 2006, for the role of metacognition in listening comprehension; Gu, 2005, in vocabulary learning; Nakatani & Goh, 2007, in oral communication; Manchón, de Larios, & Murphy, 2007, in writing). In the field of reading, many studies have shown that metacognition is closely related to reading comprehension (see A. L. Brown, 1980; Carrell, 1989; Paris & Jacobs, 1984; Paris, Lipson, & Wixson, 1983; Phakiti, 2003; Sheorey & Mokhtari, 2001; Zhang, 2010). For example, A. L. Brown (1980) postulated that readers' metacognition is closely related to their reading performance. Paris and Jacobs's (1984) study revealed a significant relationship between children's reading awareness (i.e., metacognition) and comprehension skills. Carrell's (1989) research showed close relationship between readers' metacognitive awareness and their reading ability in both their first language (L1) and second language (L2). Zhang (2010) reported that Chinese college students' metacognitive awareness was linked to their reading proficiency. The general conclusion is that skilled readers are distinguished from unskilled readers by their conscious awareness of strategic reading processes and their actual use of reading strategies. Furthermore, according to Pressley and Afflerbach (1995), reading comprehension comprises five phases: initial reading of the text, identifying important information, inference making, integrating different parts of the text, and interpreting. This framework provides the primary theoretical basis for the strategy use questionnaire in this study.

In addition, metacognitive and cognitive strategies have attracted great attention from language researchers of various areas as they reflect language learners' major strategic behaviours (e.g., Bachman & Palmer, 2010; O'Malley & Chamot, 1990; Oxford, 1990; Vandergrift et al., 2006). For example, in Bachman and Palmer's (2010) framework, language users'/test takers' metacognitive strategies are the core of strategic competence that "provide a management function in language use" (p. 48), whereas cognitive strategies are used when language users implement plans in actual language use. For the purpose of this study, metacognitive strategy use refers to test takers' conscious and purposeful mental activities that control and manage their test-taking and reading processes (Cohen & Upton, 2006; Paris & Winograd, 1990). Following established theories (e.g., Paris & Winograd, 1990; Wenden, 1998), metacognitive strategies comprise *planning* (for achieving pre-established goals), *evaluating* (for assessing tasks and personal cognitive abilities), *monitoring* (for checking and regulating performance) strategies (O'Malley & Chamot, 1990; Paris & Winograd, 1990; Wenden, 1998). Cognitive strategies in this study are viewed as specific and conscious mental behaviours or activities used by test takers to solve the problems encountered in the process of reading comprehension. They are composed of *initial reading* (for engaging in general reading of the text), *identifying important information* (for refining understanding of the text), *inference making* (for bridging information gaps in the text), and *integrating*[1] (for manipulating the text to fit information across the text) strategies (Afflerbach, 1990; Kintsch & van Dijk, 1978; Pressley & Afflerbach, 1995).

In contrast to the concensual view on the importance of strategy use in language activities, a review of literature shows that no consistent picture has emerged regarding the issue of the relationships between metacognitive and cognitive strategy use. For example, according to O'Malley and Chamot (1990), learning strategies include three types: metacognitive, cognitive, and socio-affective strategies. In addition, Oxford (1990) argued that learning strategies comprise six types: memory, cognitive, compensation, metacognitive, affective, and social strategies. According to these frameworks of strategy use, metacognitive and cognitive strategies are parallel and separate components of learning strategies. On the other hand, some researchers (e.g., Baker, 1991; Chapelle, Grabe, & Berns, 1997; Paris et al., 1991; Vandergrift et al., 2006) argued that it is hard to demarcate metacognitive and cognitive strategies, especially "when they are embedded in complex sequences of behaviour or hierarchies of decisions" (Paris et al., 1991, p. 610). In summary, the review of literature shows that not enough empirical studies have been conducted to investigate how learners' metacognitive and cognitive strategies are related in language use situations although research in this area can shed light on language users' actual processes in engaging in language tasks, indicating a gap that research should be designed to fill.

Research on Metacognitive and Cognitive Strategy Use in Language Assessment

As a reflection of the processes test takers go through in taking the test, strategy use on language tests plays an important role in validating tests and enhancing test performance (Cohen, 2006). Therefore, language testing researchers have long been interested in exploring how test takers' strategy use is related to their test performance. For example, Cohen (2006) argued that

[1]Four factors emerged in the exploratory factor analysis (EFA), though the questionnaire was designed to include five factors based on Pressley and Afflerbach's (1995) model.

when completing language tests, test takers have to “deal with both language issues and the item-response demands” (p. 308). As such, he argued that three types of strategies are involved in language tests: language learner strategies, test-management strategies, and test-wiseness strategies. Language learner strategies deal with language issues in the tests. These strategies are equivalent to reading strategies in the current study. Test-management strategies provide meaningful responses to test tasks and items, and test-wiseness strategies[2] refer to “using knowledge of testing formats and other peripheral information to obtain responses” (Cohen, 2013, p. 4).

In addition, Purpura (1997) examined the relationships between test takers’ metacognitive and cognitive strategy use in non-test contexts and their test performance on the University of Cambridge First Certificate of English Anchor Test with 1,382 English as a Foreign Language (EFL) participants using an SEM approach. Grounded in human information-processing theory (Gagnè, Yekovich, & Yekovich, 1993), his 40-item cognitive strategy use questionnaire included 11 strategy-type variables representing three underlying processing variables: comprehending, memory, and retrieval strategies. The 40-item metacognitive strategy use questionnaire involved four strategy-type variables, which represented two underlying process type variables (i.e., online and postassessment processes). As shown in his study, two underlying factors explained the second language test performance: reading ability and grammar ability. Purpura’s (1997) study showed that both metacognitive and cognitive strategy use had no direct effect on language performance, but the former was closely related to the latter. Purpura’s (1997) study is actually one of the first to investigate the relationship between grammar ability and reading ability specifically. Further, its effect was generalized across two proficiency level groups (Purpura, 1998).

X. Song (2005) abridged Purpura’s (1999) questionnaire and administered it to 161 test takers taking the Michigan English Language Assessment Battery. The result showed that the Michigan English Language Assessment Battery test takers’ use of metacognitive strategies fell into three categories: evaluating, monitoring, and assessing. X. Song and Cheng (2006) used another condensed version of Purpura’s questionnaire and assigned it to 121 Chinese college test takers taking the College English Test Band 4 (CET-4). Their study showed that CET-4 test takers used more metacognitive strategies than cognitive strategies. Both studies analyzed the relationship between strategy use and test performance by means of regression.

Phakiti (2003) developed a 35-item questionnaire including items similar to Purpura’s (1999) to examine the relationship between 384 Thai test takers’ strategy use in test contexts and their reading test performance using a multivariate analysis of variance. The cognitive strategy questionnaire focused on two factors—comprehending and retrieval—whereas the metacognitive strategies questionnaire had two factors—planning and monitoring. His study concluded that the use of cognitive and metacognitive strategies had a positive but weak relationship with reading performance, which explained 15–22% of the test score variance.

Later, based on human information-processing theory (Gagnè et al., 1993), Phakiti (2008) developed a 30-item strategy questionnaire and validated Bachman and Palmer’s (1996) theory of strategic competence with 561 Thai university students who took an EFL reading test. The results were analysed using SEM. His 17-item metacognitive strategies questionnaire comprised three subscales—Planning, Monitoring, and Evaluating strategies—whereas the 13-item cognitive strategy use questionnaire included three subscales—Comprehending, Memory, and Retrieval Strategies. The two underlying factors of the EFL reading test were lexico-grammatical

[2]Test-wiseness strategies are not within the scope of this study.

reading ability and text comprehension reading ability. Similar to Purpura (1997), his results showed that metacognitive strategy use was closely related to cognitive strategy use. In addition, lexico-grammatical ability, a better predictor of reading comprehension ability, affected text comprehension ability directly and significantly.

To sum up, previous research found that language users employed metacognitive and cognitive strategies in test contexts and both types of strategies had effects on their test performance. However, no conclusive results have been drawn regarding how metacognitive strategy use is related to cognitive strategy use as well as the effects of strategy use on test performance. This suggests that further research should be conducted in this area.

Cross-Validation With Multi-Sample Analysis

In cross-validation studies, a series of competing models derived from theories and empirical studies are tested against two (or more) samples to identify a baseline model. When the baseline model is established, tests of invariance are conducted simultaneously across these samples (Byrne, Baron, & Balev, 1996; In'nami & Koizumi, 2011; M.-Y. Song, 2011). In the area of language testing, several studies have tested factorial invariance with a multi-sample/multi-group analysis (e.g., Bae & Bachman, 1998; Ginther & Stevens, 1998; In'nami & Koizumi, 2011; Purpura, 1998; Shin, 2005; M.-Y. Song, 2011; Sticker, Rock, & Lee, 2005). Among these studies, only Purpura (1998) investigated how the relationship between metacognitive and cognitive strategy use and second language test performance varied among high and low proficiency groups in the First Certificate of English Anchor test. Thus, it shows that not enough research has been carried out to study the relationship between strategy use and test performance using this approach, which reveals a void that further research needs to fill.

Relevance to the Current Study

Findings from the literature review have the following implications for the design of the study. First, although researchers have come to a consensus that metacognition plays an important role in language use, how metacognitive strategies are related to cognitive strategies in language use situations is still not clear. Some researchers have pointed out that the distinction between metacognitive and cognitive strategies is fuzzy (Baker, 1991; Chapelle et al., 1997; Paris et al, 1991). In other words, metacognition is unitary in that it is hard to separate metacognitive and cognitive strategies when they are used in the situations in which a complicated array of decisions has to be made. Other researchers have argued that metacognitive and cognitive strategies are key components of language learners' metacognitive awareness, that is, metacognition is componential and separable (e.g., O'Malley & Chamot, 1990; Oxford, 1990; Wenden, 1998). Yet other researchers have demonstrated that metacognitive and cognitive strategy use is closely related to each other (e.g., Phakiti, 2003, 2008; Purpura, 1997, 1998). In addition, Purpura (1999) and Phakiti (2003) all raised the issue of the construct of test takers' metacognition. Phakiti (2003) argued that an important task for language testing researchers is to "measure the defined strategy construct" (p. 47). Therefore, on the basis of the existing literature, we hypothesized three models (i.e., unitary, hierarchical, and correlated models) to examine the underlying structure of metacognition and its effect on test takers' reading test performance in test contexts (see Figures 1, 2, and 3 for graphic demonstration of the models).

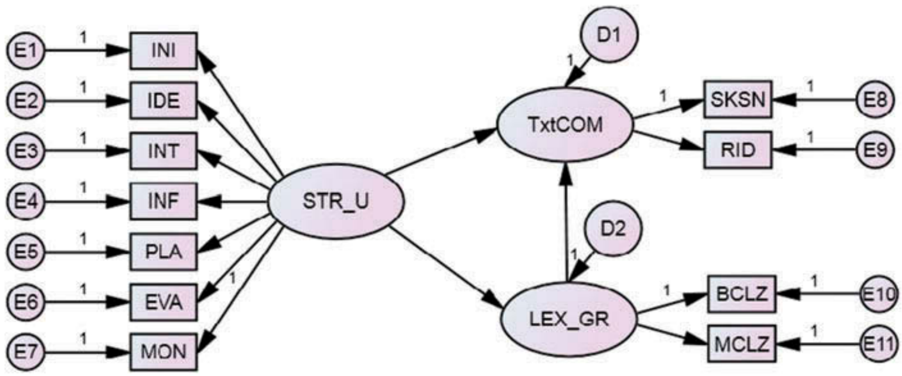

FIGURE 1 The unitary model. *Note*. INI = initial reading strategies; IDE = identifying important information strategies; INT = integrating strategies; INF = inference-making strategies; PLA = planning strategies; EVA = evaluating strategies; MON = monitoring strategies; STR_U = strategy use; TxtCOM = text comprehension reading ability; LEX-GR = lexico-grammatical reading ability; SKSN = Skimming and Scanning; RID = Reading in Depth; BCLZ = Banked Cloze; MCLZ = Multiple-Choice Cloze. (Color figure available online.)

We hypothesized that in the unitary model (Figure 1) test takers' metacognitive and cognitive strategies play a unitary role in enhancing their reading test performance. In other words, metacognitive and cognitive strategies work in synergy in affecting test performance. In the higher order model (Figure 2), test takers' strategy use was hypothesized to be hierarchical in that strategy use is a higher order factor, whereas metacognitive and cognitive strategy use are lower order factors. In the correlated model (Figure 3), test takers' metacognitive strategy use was hypothesized to correlate with their cognitive strategy use. In addition, in all these three models, strategy use (or metacognitive and cognitive strategy use) was hypothesized to have direct effect on test takers' test performance.

Second, considerable research on reading comprehension has shown that comprehension cannot occur without successful operation of lower level processes such as word recognition, syntactic parsing, and semantic-proposition encoding (e.g., Gough & Tunmer, 1986; Grabe, 2009; LaBerge & Samuels, 1974). The lower level processing knowledge is generally termed lexico-grammatical knowledge (Celce-Murcia & Larsen-Freeman, 1999; Purpura, 2004). Lexico-grammatical ability is directly related to L2 reading ability. That is, test takers' knowledge of word recognition and syntactic parsing is expected to directly affect their reading ability greatly (see Phakiti, 2008; Zhang, in press; Zhang & Zhang, 2013). Therefore, we hypothesized that the EFL reading test performance had two underlying factors: text comprehension reading ability (TxtCOM) and lexico-grammatical reading ability (LEX-GR). The latter (e.g., LEX-GR) had a direct effect on the former (e.g., TxtCOM).

In addition, according to the test syllabus of the CET-4 (National College English Testing Committee, 2006), specific skills assessed in the reading test include (a) ability to distinguish and understand the main idea and important details, and (b) ability to understand the passage by means of word knowledge. The former is represented by passage comprehension items in Skimming and

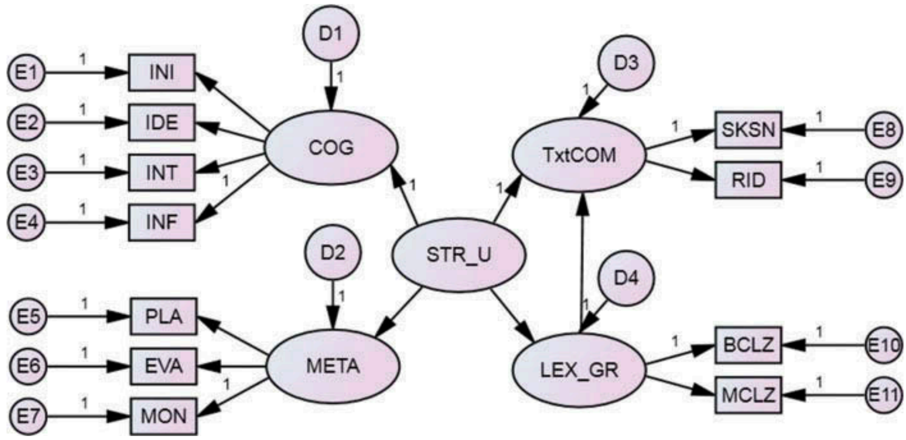

FIGURE 2 The higher-order model. *Note*. INI = initial reading strategies; IDE = identifying important information strategies; INT = integrating strategies; INF = inference-making strategies; PLA = planning strategies; EVA = evaluating strategies; MON = monitoring strategies; STR_U = strategy use; TxtCOM = text comprehension reading ability; LEX-GR = lexico-grammatical reading ability; SKSN = Skimming and Scanning; RID = Reading in Depth; BCLZ = Banked Cloze; MCLZ = Multiple-Choice Cloze. (Color figure available online.)

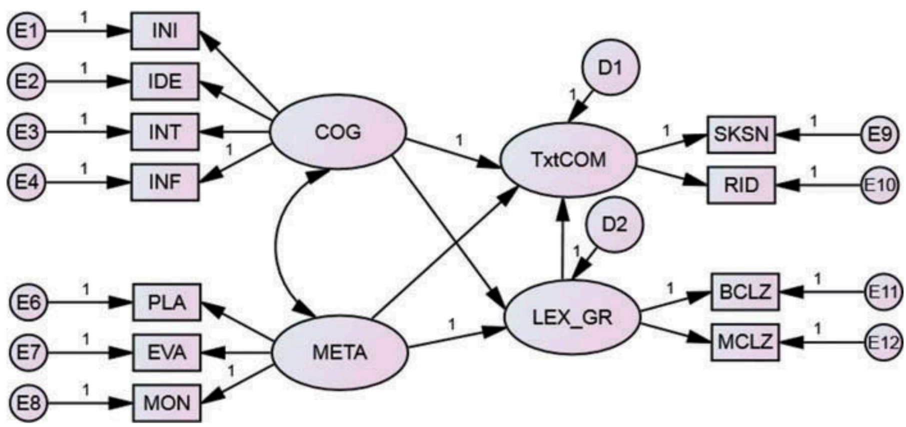

FIGURE 3 The correlated model. *Note*. INI = initial reading strategies; IDE = identifying important information strategies; INT = integrating strategies; INF = inference-making strategies; PLA = planning strategies; EVA = evaluating strategies; MON = monitoring strategies; STR_U = strategy use; TxtCOM = text comprehension reading ability; LEX-GR = lexico-grammatical reading ability; SKSN = Skimming and Scanning; RID = Reading in Depth; BCLZ = Banked Cloze; MCLZ = Multiple-Choice Cloze. (Color figure available online.)

Scanning (SKSN) and Reading in Depth (RID) sections, whereas the latter is operationalized by cloze items in Banked Cloze (BCLZ) and Multiple-Choice Cloze (MCLZ) sections. In addition, contrary to earlier assertion that cloze tests measured higher order processing abilities (Hinofotis, 1980; Oller, 1979), more recent studies have shown that cloze tests serve as a measure of lower order proficiency such as grammar and vocabulary (Alderson, 1979; Markham, 1985; Purpura, 1999, 2004; Saito, 2003; Shanahan, Kamil, & Tobin, 1982). Therefore, we hypothesized that LEX-GR was measured by test takers' performance on the BCLZ and MCLZ sections of the test and TxtCOM by the SKSN and RID.

Third, although SEM has been applied in the investigation of the relationship between test takers' reading strategy use and their test performance, to date no studies have employed multi-sample SEM to test invariance of the factor structure of the relationship between reading strategy use and reading test performance across samples of similar characteristics. Therefore, it will be interesting to examine whether the factor structure of the relationship between Chinese test takers' reading strategy use and reading test performance is generalizable across samples.

The current study addresses the following two research questions:

RQ1: What is the relationship between test takers' metacognitive and cognitive strategy use? In other words, of the three models—unitary, higher order and correlated—which model of strategy use and reading test performance fits the data best?

RQ2: What is the relationship between test takers' metacognitive and cognitive strategy use and their reading test performance? In other words, is the factor structure of the relationship between test takers' reading strategy use and reading test performance generalizable across samples?

METHOD

Settings and Participants

The participants in the current study were first year undergraduate students of non-English majors from three main types of universities in the northern part of mainland China: the arts-oriented, science-oriented, and comprehensive universities, which enroll students nationwide. For these students, English was a compulsory course in the first two years of their four year undergraduate programs.

A total of 593 Chinese college students participated in the study by filling out the consent form, answering the questionnaire, and sitting for the reading comprehension test. There were 274 (46.2 %) male and 311 (52.4 %) female students between the ages of 18 and 24 ($M = 19.37$, $SD = 0.98$). Eight other students (1.4 %) did not indicate their gender. On average, they had received 9.19 years ($SD = 2.41$) of formal English instruction by the time of the study.

Instruments

Two instruments were used in the study: the Metacognitive and Cognitive Strategy Questionnaire and the CET-4 Reading subtest.

The Metacognitive and Cognitive Strategy Questionnaire. The metacognitive strategy questionnaire was based on the theory of metacognition (e.g., Paris & Winograd, 1990; Wenden, 1998) and Cohen and Upton's (2006) framework, whereas the cognitive strategy questionnaire was grounded in Pressley and Afflerbach's (1995) constructively responsive reading model. Strategy use items were selected from the literature on learning strategies (e.g., O'Malley & Chamot, 1990; Oxford, 1990; Purpura, 1999), reading strategies (e.g., Carrell, 1989; Mohktari & Reichard, 2002; Phakiti, 2003, 2008; Pressley & Afflerbach, 1995; Sheorey & Mohktari, 2001), and test-taking studies (e.g., Anderson, 1991; Anderson, Bachman, Perskins, & Cohen, 1991; Cohen & Upton, 2006). The questionnaire used a 6-point Likert scale: 0 (*never*), 1 (*rarely*), 2 (*sometimes*), 3 (*often*), 4 (*usually*), and 5 (*always*). One expert in L1 reading and two experts in L2 reading and testing reviewed the pool of items evaluating its content validity, clarity, readability, and redundancy. The questionnaire was then piloted with a group of students ($N = 71$) to identify the ambiguous or confusing items with regard to wording, format, and content. Next, 650 second-year undergraduate students between ages 18 and 24 ($M = 20.58$, $SD = 1.21$) from the same three universities were invited to respond to the questionnaire before it was used in the main study. An EFA was performed to explore the clustering of items and identify the potential subscales of the questionnaire. Seven factors were generated, accounting for 45.26% of the total variance. We then decided that items loaded on more than one factor were to be deleted out of consideration for simplicity in structure (J. D. Brown, 2001; Dörnyei, 2003). Factor loadings greater than .30 were reported. This led to a total of 38 items measuring seven subscales: Planning, Evaluating, Monitoring, Initial Reading, Identifying Important Information, Integrating, and Inference-Making strategies (see Table 1). The questionnaire is presented in Appendix A.

To validate the questionnaire with the 593 participants, an analysis of factorial structure was conducted at the item level. The posited confirmatory factor analysis (CFA) model showed acceptable model fit, $\chi^2(590) = 1000.64$, χ^2/df ratio= 1.70, root mean square error of approximation (RMSEA) = .034, normed fit index (NFI) = .90, comparative fit index (CFI) = .91, standardized root mean square residual (SRMR) = .047. On the basis of this result, we generated composite variables at the subscale level of the questionnaire for further SEM analyses. This technique is termed item-parceling (Bandalos & Finney, 2001; Little, Cunningham, Shahar, & Widaman, 2002), which is commonly used in latent variable analyses (Kunnan, 1998; Purpura, 1999). A parcel is an aggregate-level indicator, comprising the sum or average of two or more

TABLE 1
Subscales of the Metacognitive and Cognitive Strategy Use Questionnaire

Strategy	*No. of Items*	*Items Used*	*Reliability (Cronbach's α)*
Planning (PLA)	6	1, 2, 3, 4, 5, 6	.620
Evaluating (EVA)	8	7, 8, 9, 10, 11, 12, 13, 14	.836
Monitoring (MON)	10	15, 16, 17, 18, 19, 20, 21, 22, 23, 24	.785
Initial reading (INI)	3	25, 26, 27	.486
Identifying important information (IDE)	4	28, 29, 30, 31	.564
Integrating (INT)	4	32, 33, 34, 35	.701
Inference making (INF)	3	36, 37, 38	.665
Total	38		.888

TABLE 2
Subsections of the Reading Comprehension Test With Reliability Estimates

Section	*No. of Items*	*Items Used*	*Reliability (Cronbach's α)*
Skimming and Scanning (SKSN)	10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	.686
Banked Cloze (BCLZ)	10	11, 12, 13, 14, 15, 16, 17, 18, 19, 20	.742
Reading in Depth (RID)	9	21 22, 23, 24, 25, 26, 27, 29, 30	.569
Multiple-Choice Cloze (MCLZ)	19	31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 50	.876
Total	48		.903

items. These parcels were used for analyses in the main study. The number of items used for the variables in the questionnaire and the test are presented in Tables 1 and 2.

The CET-4 Reading subtest. A commercially published version of the CET-4 Reading subtest (Fang, 2010) was adopted to measure test takers' reading test performance. As one of the most influential college tests in China (Jin, 2008), the CET is administered by the National College English Testing Committee on behalf of the Chinese Ministry of Education (see Zheng & Cheng, 2008).[3] The CET is a test battery that includes the CET-4, the CET-6, and the CET-Spoken English Test. As a nationwide standardized test, it has been subjected to rigorous validation processes to ensure its high quality as an assessment tool (Yang & Weir, 1998). The CET-4 Reading test in this study comprises 50 items, including four sections: 10 SKSN items, 10 BCLZ items, 10 items measuring RID, and 20 MCLZ items. The sample items of the reading test are presented in Appendix B. Prior to the main study, we did a preliminary analysis of the test by performing EFA with the first half of the sample ($N = 296$) and CFA with the second half ($N = 297$) for cross-validation (Bollen, 1989).

A matrix of tetrachoric correlation using all 50 items was generated in PRELIS2 and exported into IBM SPSS Statistics Version 20 for further analysis. A series of EFA performed on each section of the test. The results of EFA showed that Items 28 and 42 had very low loadings on the extracted factors. After examining the items carefully, we decided to drop them in later analysis as they might not tap into the required skills (Purpura, 1999). The subsequent CFA produced acceptable model fitness, $\chi^2(2) = 6.159$, χ^2/df ratio = 3.079, RMSEA = .084, NFI = .96, CFI = .99, SRMR = .020. On the basis of the CFA results, composite variables at the subsection level of the test were generated for the main study.

Data Collection, Preparation, and Analysis

Data collection and preparation. Before being administered to the participants, the consent form and the questionnaire were translated into Chinese, followed by the back-translation procedure. Test takers were required to complete the questionnaire within 30 to 40 min, and the CET-4 Reading subtest within 55 min. All test items were scored dichotomously. We marked and double-checked every answer to ensure all the items were scored and entered into the database accurately.

[3]Although for the sake of clarity, the term CET-4 is used in this study, the data collected were from a commercially published CET-4 study guide.

Preliminary statistical analyses. To conduct multi-sample SEM analyses, the data were randomly split into two halves ($N = 296$ for Sample 1 and $N = 297$ for Sample 2; MacCallum, Roznowski, Mar, & Reith, 1994). Descriptive statistics and reliability at the item and subscale level of the questionnaire and subsection level of the reading test were calculated for each sample. Assumptions regarding univariate normality and multivariate normality were also inspected. Values of skewness within ± 3 and kurtosis within ± 10 indicated univariate normality (Kline, 2011). Multivariate normality was evaluated using Mardia's coefficient and a value of 5.00 or below represented multivariate normality (Byrne, 2006).

SEM. Prior to conducting multi-sample analyses, we tested the three hypothesized models for each sample separately to identify a baseline model (In'nami & Koizumi, 2011; M.-Y. Song, 2011). After the baseline model was selected, cross-group invariance was tested by placing constraints on sets of parameters in a logically ordered and increasing restrictive manner (Byrne, 2011). Figures 1 to 3 presents the three hypothesized models of strategy use and reading test performance tested: (a) a unitary model (Figure 1), (b) a higher order model (Figure 2), and (c) a correlated model (Figure 3).

Multiple fit indices were calculated to investigate the fit of the model. The non-significant value of chi-square indicates good model fit. However, because it is sensitive to sample size (Kline, 2011), the chi-square to degree of freedom ratio is normally calculated and a value less than 3 is considered to indicate a well-fitting model. In addition, the absolute fit indices were calculated. The RMSEA shows how well a model fits the population and should be less than .08 to indicate reasonable error of approximation (Browne & Cudeck, 1993). A narrower RMSEA 90% confidence interval is indicative of better model fit (Kline, 2011). The SRMR evaluates the differences between observed and predicted variance and covariance. Values below .10 indicate a good model fit (Kline, 2011). The lower values of the Akaike Information Criteria and the Consistent Akaike Information Criteria also indicate good model fit. A chi-square difference test was used to compare models.

Multi-Sample SEM. After the best-fitting model among the three hypothesized models was selected for both samples, we performed a multi-sample SEM analysis to cross-validate the selected model to test the invariance of factor loadings, measurement error variances, structural regression coefficients, and factor variances of the baseline model. Figure 4 displays the statistical procedures followed in this study.

We used IBM SPSS AMOS computer program, Version 20.0 (Arbuckle, 2011) to perform the analysis. Maximum Likelihood technique was chosen as the method of parameter estimation.

RESULTS

Descriptive Statistics

Descriptive statistics at the item level of the questionnaire and reading test were first calculated. We then calculated the descriptive statistics at the subscale level of the questionnaire and subsection level of the test (see Table 3). All values of skewness and kurtosis were within the accepted range for univariate normality. Multivariate normality was represented by a Mardia's

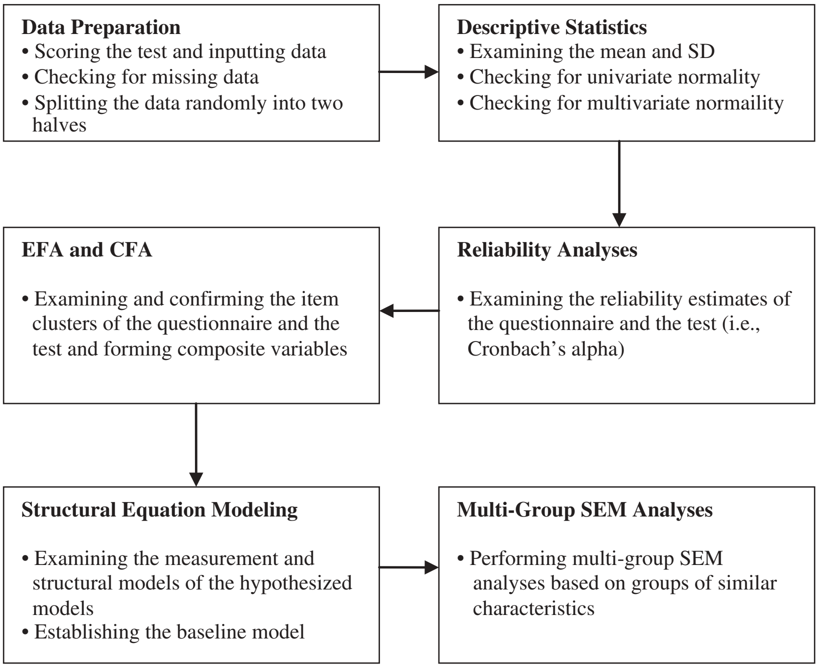

FIGURE 4 A flow chart of the statistical procedures used in this study (based on Purpura, 1999). *Note.* EFA = exploratory factor analysis; CFA = confirmatory factor analysis; SEM = structural equation model.

coefficient smaller than 5.00, with 3.136 for Sample 1 and 3.605 for Sample 2. Reliability estimates for the subscales of the questionnaire and the subsections of the test are shown in Table 1 and 2. Reliability estimates for the questionnaire scale and the reading test were .888 and .903, respectively.

SEM

First, to establish the baseline model, we tested the three hypothesized models with both samples. As shown in Table 4, *the unitary model* fit the data well. In spite of the fact that the chi-square statistic was significant (i.e., $\chi^2(43) = 109.74$, $p < .05$, for Sample 1; $\chi^2(43) = 67.53$, $p < .05$, for Sample 2), the other fit indices showed a good model fit with the data: CFI = .92, RMSEA = .073, 90% confidence interval (CI) [.056, .089], and SRMR = .057 for Sample 1; CFI = .97, RMSEA = .044, 90% CI [.022, .063], and SRMR = .044 for Sample 2. The standardized direct effects and error/disturbance of the unitary model are presented in Appendix C.

The higher-order model also seemed to fit the data well, but it had a negative error variance for the metacognitive strategy factor. If the problematic variance is fixed to zero to solve the problem, the model becomes meaningless and not interpretable.

The correlated model also had the similar problem of a negative error variance associated with RID and MCLZ. In addition, it showed poor model fit across samples.

TABLE 3
Descriptive Statistics for Sample 1 and Sample 2 of the Questionnaire and the CET-4 Reading Subtest

	M		*SD*		*Skewness*		*Kurtosis*	
	Sample 1	*Sample 2*	*Sample 1*	*Sample 2*	*Sample 1*	*Sample 2*	*Sample 1*	*Sample 2*
Strategy Questionnaire								
Initial reading (INI)	3.09	3.12	.67	.78	.01	−.01	.39	−.23
Identifying important information (IDE)	3.41	3.41	.68	.74	−.13	−.23	.02	−.40
Integrating (INT)	3.75	3.82	.72	.75	−.28	−.36	−.40	−.30
Inference-making (INF)	3.52	3.47	.85	.86	−.35	−.28	−.29	−.26
Planning (PLA)	3.23	3.25	.84	.91	−.26	−.32	−.13	−.01
Evaluating (EVA)	2.71	2.73	.69	.73	−.17	.11	−.01	−.19
Monitoring (MON)	3.56	3.56	.59	.66	−.03	−.27	−.25	−.20
CET-4 Reading subtest								
Skimming and Scanning (SKSN)	6.61	6.68	2.05	2.11	−.77	−.99	.10	.53
Banked Cloze (BCLZ)	2.72	2.80	1.31	1.35	−.18	−.22	−.77	−.83
Reading in Depth (RID)	11.74	12.86	3.60	3.87	−.82	−.65	.20	.29
Multiple-Choice Cloze (MCLZ)	5.33	5.64	2.62	2.60	−.61	−.62	−.65	−.51

Note. CET-4 = College English Test Band 4. Sample 1: $N = 296$; Sample 2: $N = 297$.

TABLE 4
Fit Indices for the Three Models With the Two Samples

	χ^2	*df*	χ^2/df	*CFI*	*RMSEA*	*RMSEA 90% CI*	*AIC*	*CAIC*	*SRMR*
Sample 1									
Unitary model	109.74*	43	2.55	.92	.073	[.056, .089]	177.74	278.06	.057
Higher-order model	111.31*	40	2.78	.91	.078	[.061, .095]	185.31	300.85	.071
Correlated model	202.58*	41	4.94	.80	.116	[.100, .132]	274.58	368.66	NA
Sample 2									
Unitary model	67.53*	43	1.57	.97	.044	[.022, .063]	135.53	228.45	.044
Higher-order model	81.55*	40	2.04	.96	.059	[.041, .078]	155.55	259.25	.059
Correlated model	165.64*	41	4.04	.87	.101	[.086, .118]	237.64	337.43	.101

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; AIC = Akaike Information Criteria; CAIC = Consistent Akaike Information Criteria; SRMR = standardized root mean square residual.
$^*p < .05$.

Based on these results, the unitary model was selected as the baseline model that fit the data well both statistically and substantively. Thus, the unitary model will be used in cross-validation analyses.

Multi-Sample SEM Analysis for Cross-Validation

In the multi-sample SEM analysis, the unitary model was tested across samples (a) with no constraints applied; (b) with factor loading constrained; (c) with factor loadings and error variance constrained; (d) with factor loadings, error variance, and structural regression coefficients constrained; and (e) with factor loadings, error variance, structural regression coefficients, and factor variance constrained. The test was conducted in an increasingly restrictive manner with the most stringent constraints in the last model (Model 5). First, we tested the baseline model (i.e., the unitary model) across two samples with no equality constraints. As shown in Table 5, the fit indices showed that this model fit the data well with both samples: CFI = .951, RMESA = .042, 90% CI [.033, .048], SRMR = .0568.

Second, we tested the invariance of factor loadings by placing constraints on factor loadings with both samples, meaning constraining all factor loadings across samples to be equal. This was more stringent compared with the first step which released no constraints. As indicated in Table 5, Model 2 yielded good fit indices: CFI = .950, RMSEA = .041, 90% CI [.032, .049], SRMR = .0555.

Third, we placed constraints on factor loadings and error variances across the samples to test the invariance of these parameters. As a result, Model 3 produced good model fit to the data: CFI = .948, RMSEA = .039, 90% CI [.031, .047], SRMR = .0563 (see Table 5).

Fourth, we took a more stringent step to test the invariance of the factor loadings, error variances, and structural regression coefficients by constraining all these parameters across both samples. As shown in Table 5, the fit indices of Model 4 showed that this model fit the data well: CFI = .949, RMSEA = .038, 90% CI [.030, .047], SRMR = .0610.

Finally, the invariance of the factor loadings, error variances, structural regression coefficients, and factor variances was tested. We placed constraints on all these parameters, which is the most stringent level of the invariance test. As shown in Table 5, Model 5 fit the data well: CFI = .948, RMSEA = .038, 90% CI [.030, .046], SRMR = .0616, indicating all the factor loadings, error variance, structural regression coefficients, and factor variance were equal across the sample.

As previously discussed and shown in Table 5, all five tested models fit the data well. Because Model 2, Model 3, Model 4, and Model 5 all nested within Model 1, we then conducted the chi-square difference tests to examine if the four models were significantly different from Model 1. As shown in Table 6, the chi-square difference test indicated that all the four models were not

TABLE 5
Fit Indices for the Unitary Model for Cross-Validation

	χ^2	*df*	*CFI*	*RMSEA*	*RMSEA 90%CI*	*AIC*	*SRMR*
Model 1: Baseline: no equality constraints	169.018*	82	.951	.042	[.033, .051]	331.018	.0568
Model 2: Factor loadings equal	178.041*	90	.950	.041	[.032, .049]	306.041	.0555
Model 3: Factor loadings and error variance equal	192.250*	101	.948	.039	[.031, .047]	298.250	.0563
Model 4: Factor loadings, error variance, and structural regression coefficients equal	194.691*	104	.949	.038	[.030, .047]	294.691	.0610
Model 5: Factor loadings, error variance, structural regression coefficients, and factor variance equal	199.732*	107	.948	.038	[.030, .046]	293.732	.0616

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; AIC = Akaike Information Criteria; SRMR = standardized root mean square residual.
$^*p < .05$.

TABLE 6
Chi-Square Difference Test Results

	$\Delta\ \chi^2$	$\Delta\ df$	*p*	*Significance*
Model 1 vs. Model 2	9.023	8	.340	*ns*
Model 1 vs. Model 3	14.209	11	.222	*ns*
Model 1 vs. Model 4	1.841	3	.606	*ns*
Model 1 vs. Model 5	5.641	3	.130	*ns*

Note. ns = not significant

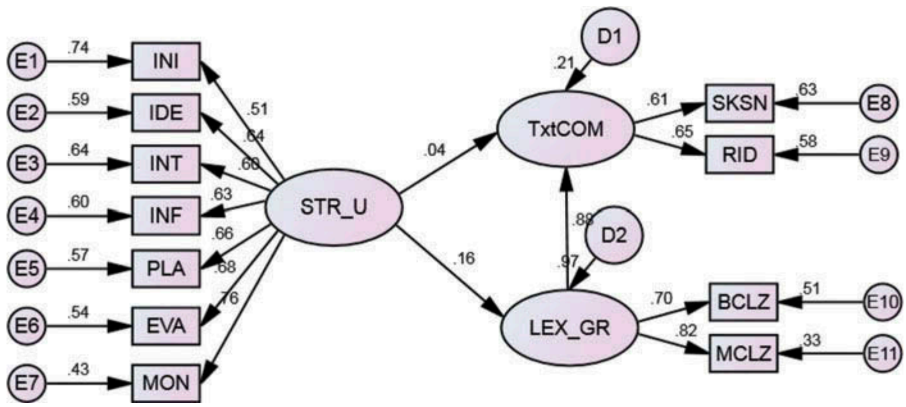

FIGURE 5 The final structural equation model with standardized estimates (*N*s = 296, 297). *Note*. INI = initial reading strategies; IDE = identifying important information strategies; INT = integrating strategies; INF = inference-making strategies; PLA = planning strategies; EVA = evaluating strategies; MON = monitoring strategies; STR_U = strategy use; TxtCOM = text comprehension reading ability; LEX-GR = lexico-grammatical reading ability; SKSN = Skimming and Scanning; RID = Reading in Depth; BCLZ = Banked Cloze; MCLZ = Multiple-Choice Cloze. (Color figure available online.)

significantly different from Model 1, suggesting that the invariance of factor loadings, error variance, structural regression coefficients, and factor variance was supported across Sample 1 and Sample 2. The final tested model, which is discussed further in following sections, is presented in Figure 5.

DISCUSSION

In this study, we have examined Chinese college test takers' metacognitive and cognitive strategy use and their reading test performance through multi-sample SEM analysis. This section discusses the results in relation to the two research questions.

RQ1: What is the relationship between test takers' metacognitive and cognitive strategy use? In other words, of the three models—unitary, higher order, and correlated—which model of strategy use and reading test performance fit the data best?

On the basis of relevant literature, the unitary, higher order, and correlated models were hypothesized, tested, and compared to identify the baseline model. Our analyses showed that the unitary model proved to be the best-fitting and the baseline model for the cross-validation study. Although the higher order model also yielded a good model fit, we decided not to select it due to the issue of negative error variance. In other words, if we start to solve the problem statistically, the model will be meaningless. Therefore, the unitary model was selected as the baseline model for the cross-validation study.

With regard to the functions of metacognitive and cognitive strategies and how they are related in language use, scholars have provided taxonomies related to the nature of these strategies, suggesting that it is possible in theory to distinguish different types of strategies within the overarching construct of strategy use. For example, O'Malley and Chamot (1990) classified learner strategies into three types—metacognitive, cognitive, and socio-affective—whereas Oxford (1990) divided learning strategies into six kinds—memory, cognitive, compensation, metacognitive, affective, and social strategies—in her Strategy Inventory for Language Learners. Our analyses showed that the higher order model did not fit the data well, suggesting that metacognitive and cognitive strategies may not be so clearly distinguishable when in use.

We hypothesized three models of strategy use (i.e., unitary model, higher order model, and correlated model) to investigate the relationships between metacognitive, cognitive strategy use, and reading test performance. In postulating these models, we are concerned with language use strategies rather than language learning strategies. As argued by Cohen (1998), language learner strategies are generally categorized into two types: language learning and language use strategies. Language learning strategies are general strategies that are purposefully employed by language learners to continuously enhance their language learning; by contrast, language use strategies are specific strategies that are employed by language users to improve language performance in specific situations. In other words, the analysis results of this study shed light on the relationship between metacognitive and cognitive strategies in actual use contexts (i.e., test context).

Purpura (1997, 1998, 1999) and Phakiti (2003, 2008) concluded that metacognitive and cognitive strategies appeared to be closely related, and they all raised the issue of the relationships between metacognitive and cognitive strategy use in the test context. For example, Purpura (1999) pointed out explicitly that "cognitive strategy use seems to function in concert with metacognitive strategy use" (p. 127), indicating that test takers need to use both metacognitive and cognitive strategies simultaneously to optimize their test performance. Phakiti (2003) also found that "most cognitive strategies occurred in association with metacognitive strategies" (p. 43). Therefore, he argued that metacognitive and cognitive strategy use seemed to "form a continuum" (p. 44).

On the basis of our analysis, the good fit of the unitary model with the data lent support to the relationship between metacognitive and cognitive strategies in actual use situation, that is, the test context. The finding indicates that language users employed both metacognitive and cognitive strategies in the test context. Furthermore, language users and test takers used metacognitive and cognitive strategies collectively which function in a unitary manner. This is congruent with researchers' earlier views (e.g., Baker, 1991; Chapelle et al., 1997; Goh, 2002;

Paris et al., 1991; Zhang, Aryadoust, & Zhang, 2013) that the distinction between metacognitive and cognitive strategies hinges on the variation of topic, task, and individuals involved. For example, Paris et al.'s (1991) argued that it is difficult to make distinction between metacognitive and cognitive strategies "when they are embedded in complex sequences of behaviour or hierarchies of decisions" (p. 610). In addition, the finding backs up the previous argument that metacognitive and cognitive strategies seem to form a continuum and function in concert (Phakiti, 2008; Purpura, 1999).

This shows that when language users are faced with a series of complex behaviours or decisions, the strategies they employ to deal with the required tasks are not clearly distinguishable. In test contexts similar to the one in this study, a wide range of sources of information and task demands are presented to test takers who work under time constraints. Therefore, they tend to use multiple strategies simultaneously to deal with language and test task demands to maximize their test performance. In other words, metacognitive and cognitive strategies cannot be separated in real language use situations. This is substantiated by the unitary model in which metacognitive and cognitive strategies function in synergy and collectively explain a significant portion of the variance in reading test performance in a unitary manner. The synergy of metacognitive and cognitive strategies has also been observed in listening, the other language reception skill (Goh, 2002; Vandergrift, 2003).

Our study about strategy use in test contexts can be viewed as empirical validation of argument about the relationship between metacognitive and cognitive strategies in language use situations. For example, this study provides empirical evidence for Bachman and Palmer's (2010) revised language use model. In their recent model, cognitive strategies, a newly added component, are perceived as part of language users/test takers' peripheral attributes, whereas metacognitive strategies, the core of strategic competence, are viewed as part of test takers' focal attributes. As argued by Bachman and Palmer (1996), strategic competence links other components of individuals' characteristics. In other words, language users' metacognitive and cognitive strategy use is related to each other. However, it is still not clear how they are connected in actual language use situations, though the finding will shed light on language users' mental processes while taking tests. Our study thus serves as one of the few empirical studies that explore the relationship between metacognitive and cognitive strategies in test contexts. As shown in our analysis, test takers employed cognitive strategies as well as metacognitive strategies and the two types of strategies functioned in synergy to maximize their reading test performance. To sum up, our finding provides validating evidence for Bachman and Palmer's (2010) language use model in that it not only indicates the plausibility of adding cognitive strategies to this model but also reveals how metacognitive and cognitive strategies are related empirically.

RQ2: What is the relationship between test takers' metacognitive and cognitive strategy use and their reading test performance? In other words, is the factor structure of the relationship between test takers' reading strategy use and reading test performance generalizable across samples?

The cross-validation study showed the invariance of the factor loadings, error variances, structural regression coefficients, and factor variances across the two samples, indicating that the unitary model of test takers' strategy use and reading test performance was generalizable across samples. This showed that metacognitive and cognitive strategies appeared to play a unitary role in enhancing the Chinese college test takers' reading test performance.

Based on the final model identified (see Figure 5), we found that seven variables of metacognitive and cognitive strategy use loaded on strategy use (STR_U) with values ranging from .51 to .76, suggesting that the latent variable STR_U was well defined by the measured variables. Among the seven subscales of strategy use, three subscales of metacognitive strategy use had the highest loadings (i.e., $\beta = .76$ for monitoring strategies, .68 for evaluating strategies, and .66 for planning strategies), indicating that the STR-U was better defined by metacognitive strategies than cognitive strategies.

Regarding the factorial structure of the reading test, our findings are similar to Phakiti (2008) in that the CET-4 Reading subtest had two underlying factors: LEX-GR and TxtCOM. LEX-GR was well measured by the test takers' performance on BCLZ ($\beta = .70$) and MCLZ ($\beta = .82$) sections of the test and TxtCOM by their performance on the section of SKSM ($\beta = .61$) and RID ($\beta = .65$), suggesting that the four measured variables defined the two latent variables well. In addition, LEX-GR had a direct and significant effect on TxtCOM ($\beta = .88$), indicating that the former affected the latter greatly, but also showed that they were distinct constructs. All paths in the model were statistically significant ($p < .05$). This finding was consistent with relevant theories and empirical studies in that lexico-grammatical ability was found to affect reading comprehension ability to a great extent (see Gough & Tunmer, 1986; Grabe, 2009; LaBerge & Samuels, 1974; Phakiti, 2008; Zhang, in press; Zhang & Zhang, 2013). It also indicates that the model of the CET-4 Reading subtest identified in our study appeared to be consistent with the test syllabus of the CET-4 (National College English Testing Committee, 2006). With regard to the relationship between test takers' strategy use and reading test performance, we found that test takers' strategy use affected their LEX-GR significantly ($\beta = .16$, $p < .05$), whereas it had an indirect effect on TxtCOM ($\beta = .04$). According to Rumelhart's (2004) and Stanovich's (1980) information-processing model, readers construct meaning from the text using multiple tools, which means that they will take compensatory measures when they encounter problems. In the current scenario, test takers will use strategies to make up for their lack of proficiency. For the items tapping into test takers' lexico-grammatical reading ability, strategies played a relatively important role. However, for the items tapping into test takers' text comprehension reading ability, such as the items in the section of RID, which are assumed to measure students' reading ability at a higher level, strategy use played a minor role in compensating for their lack of proficiency. This result is congruent with Phakiti's (2008) finding that cognitive strategy use explained 16–30% of test takers' lexico-grammatical performance. In addition, this finding also concurs with Bachman's (1990) argument that strategy use is only one part of test takers' characteristics among the factors that affect performance on language tests. Language ability is still the dominant contributor to test takers' test performance.

CONCLUSIONS AND LIMITATIONS

In this study we investigated the relationship between Chinese college test takers' metacognitive and cognitive strategy use and reading test performance using the multi-sample SEM approach. Results showed that test takers' metacognitive strategies functioned in concert with cognitive strategies in a unitary manner in enhancing their reading test performance. In addition, it was found that test takers' strategy use had a significant effect on their lexico-grammatical reading ability.

Findings from this study provide empirical and validating evidence for Bachman and Palmer's (2010) updated model of language use. Although Phakiti (2008) conducted a longitudinal study validating Bachman and Palmer's (1996) strategic competence model on EFL reading tests, no studies have been carried out to examine their updated model empirically. Thus, our study is expected to make a contribution in filling this gap.

In addition, aside from Purpura's (1998) study, no reading research has been done to investigate test takers' strategy use and reading test performance using the multi-group SEM approach. Our study serves as an exploratory attempt to conduct a test takers' strategy use study using multi-sample SEM analyses across groups of similar characteristics. We hope it will attract language testing researchers' attention to using diverse methods to address the intriguing issues related to test-taking processes and test validation (Aryadoust, 2013).

Our findings have practical implications for classroom instruction in reading comprehension strategies and test-management strategies. Our study found that test takers' strategy use appears to improve test takers' reading test performance though it is limited to the items measuring lexico-grammatical knowledge, which is essential to reading comprehension. For the items assessing higher level reading ability, strategy use appears to play a less important role. This suggests that instruction on reading or test-management strategies may be limited in improving test takers' reading performance. Thus, classroom instructors would need to focus on improving students' language knowledge to support and enhance their reading ability. Furthermore, they would need to train learners to use relevant strategies employing not only contextual clues for word-level inference making but also more general discourse cues for successful higher order comprehension.

However, although this study has revealed some interesting findings, it should be stressed that due to the limitation to the sample size and geographical sites of the participants, the generalization of the results to the entire CET-4 population or to other reading tests might be restricted. Therefore, it is suggested that future research in this area be done with a larger CET-4 sample, or with other reading comprehension tests from different cultural and educational contexts and with samples of different demographical characteristics. In addition, given a larger sample size, it is recommended that future research be done to cross-validate the findings from this study with additional samples of similar characteristics (Byrne, Baron, & Balev, 1998).

Another limitation of the study concerns the employment of self-report questionnaires as a measuring tool of test takers' metacognitive and cognitive strategies. Questionnaires could be inaccurate and imprecise in measuring test takers' strategy use. For example, participants may report only the strategies they should employ instead of those they actually used. Second, participants might have found it hard to distinguish between metacognitive and cognitive strategies as they are not easily distinguishable, especially in specific use situations similar to the one in this study. Finally, the questionnaire instrument may not be the best tool to capture the complicated mental processes test takers go through in taking a test. Therefore, it is suggested that future research adopt a mixed method design, that is, utilizing a qualitative approach to complement and triangulate the findings from a quantitative study (Dörnyei, 2007). It is hoped that such an approach will provide a more thorough and complete understanding of the relationship between test takers' strategy use and test performance.

REFERENCES

Afflerbach, P. P. (1990). The influence of prior knowledge on expert readers' main idea construction strategies. *Reading Research Quarterly*, *25*, 31–46

Alderson, J. C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, *13*, 219–227.

Anderson, N. (1991). Individual differences in strategy use in second language reading and testing. *Modern Language Journal*, *75*, 460–472

Anderson, N. J., Bachman, L., Perskins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing*, *8*, 41–66.

Arbuckle, J. L. (2011). IBM SPSS Amos 20.0 [Computer software]. New York, NY: IBM

Aryadoust, V. (2013). *Building a validity argument for a listening test of academic proficiency*. Newcastle, UK: Cambridge Scholors Publishing.

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

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

Bachman, L. F., & Palmer, A. S. (2010). *Language testing in practice*. Oxford, UK: Oxford University Press.

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.

Baker, L. (1991). Metacognition, reading, and science education. In C. Santa & D. Alvermann (Eds.), *Science learning: Processes and applications* (pp. 2–13). Newark, DE: International Reading Association.

Bandalos, D. L., & Finney, S. J. (2001). *Item parceling issues in structural equation modeling*. Mahwah, NJ: Erlbaum.

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

Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. C. Bruce, & W.F. Brewer (Eds.) *Theoretical issues in reading comprehension: Perspectives from psychology, linguistics, artificial intelligence and education* (pp. 453–482). Hillsdale, NJ: Erlbaum.

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge, UK: Cambridge University Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park CA: Sage

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Erlbaum.

Byrne, B. M., Baron, P., & Balev, J. (1996). The Beck Depression Inventory: Testing for its factorial validity and invariance across gender for Bulgarian adolescents. *Personality and Individual Differences*, *21*, 641–651.

Byrne, B. M., Baron, P., & Balev, J. (1998). The Beck Depression Inventory: A cross-validated test of factorial structure for Bulgarian adolescents. *Educational and Psychological Measurement*, *58*, 241–251.

Carrell, P. L. (1989). Metacognitive awareness and second language reading. *Modern Language Journal*, *73*, 121–131.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course*. Boston, MA: Heinle and Heinle.

Chapelle, C., Grabe, W., & Berns, M. (1997). *Communicative language proficiency: Definitions and implications for TOEFL 2000* (TOEFL Monograph Series No. 10). Princeton, NJ: Educational Testing Service.

Cohen, A. D. (1998). Strategies and process in test taking and SLA. In L. F. Bachman & A. D. Cohen (Eds.), *Interface between second language acquisition and language testing research* (pp. 90–111). Cambridge, UK: Cambridge University Press.

Cohen, A. D. (2006). The coming age of research on test-taking strategies. *Language Assessment Quarterly*, *3*, 307–331.

Cohen, A. D. (2013). Using test-wiseness strategy research in task development. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 893–905). Hoboken, NJ: Wiley/Blackwell.

Cohen, A. D., & Upton, T. A. (2006). *Strategies in responding to the new TOEFL reading tasks* (Monograph No. 33). Princeton, NJ: ETS. Retrieved from http://www.ets.org/Media/Research/pdf/RR-06-06.pdf

Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration and processing*. Mahwah, NJ: Erlbaum.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford, UK: Oxford University Press.

Fang, Z. (2010). *Complete guide to the College English Test Band 4*. Beijing, China: Foreign Language Teaching and Research Press

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-inquiry. *American Psychologist*, *34*, 906–911.

Gagnѐ, E. D., Yekovich, C. W., & Yekovich, F. R. (1993). *The cognitive psychology of school learning*. New York, NY: Harper Collins.

Ginther, A., & Stevens, J. (1998). Language background, ethnicity, and the internal construct validity of the Advanced Placement Spanish Language Examination. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp. 169–194). Mahwah, NJ: Erlbaum.

Goh, C. (1998). How learners with different listening abilities use comprehension strategies and tactics. *Language Teaching Research*, *2*, 124–147.

Goh, C. (2002). Exploring listening comprehension tactics and their interaction patterns. *System*, pp. 185–206.

Goh, C. (2008). Metacognitive instruction for second language listening development: Theory, practice and research implications. *RELC Journal*, *39*, 188–213.

Gough, P. B., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*, 6–10.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge, UK: Cambridge University Press.

Grabe, W., & Stoller, F. (2002). *Teaching and researching reading*. New York, NY: Pearson Education.

Gu, P. Y. (2005). *Vocabulary learning strategies in the Chinese EFL context*. Singapore, Singapore: Marshall Cavendish International Private Limited.

Hinofotis, F. B. (1980). Cloze as an alternative method of ESL placement and proficiency testing. In J. W. Oller & K. Perkins (Eds.), *Research in language testing* (pp. 121–128). Rowley, MA: Newbury House.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternative. *Structural Equation Modeling*, *6*, 1–55.

In'nami, Y., & Koizumi, R. (2011). Factor structure of the revised TOEIC® test: A multiple-analysis. *Language Testing*, *29*, 131–152.

Jin, Y. (2008). Powerful tests, powerfulness test designers? Changes facing the College English Test. *CELEA Journal*, *31*, 3–11.

Kintsch, W., & van Dijk. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Kline, R. B. (2011). *Principles and practices of structural equation modeling* (2nd ed.). New York, NY: Guilford

Kunnan, A. J. (1995). *Test taker characteristics and test performance: A structural modeling approach*. Cambridge, UK: Cambridge University Press.

Kunnan, A. J. (1998). An introduction to structural equation modelling for language assessment research. *Language Testing*, *15*, 295–332.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing. *Cognitive Psychology*, *6*, 293–323.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*, 151–173.

MacCallum, R. C., Roznowski, M., Mar, C. M., & Reith, J. V. (1994). Alternative strategies for cross-validation of covariance structure models. *Multivariate Behavioural Research*, *29*, 1–32.

Manchón, R. M., de Larios, J. R., & Murphy, L. (2007). A review of writing strategies: Focus on conceptualizations and impact of first language. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice* (pp. 229–250). Oxford, UK: Oxford University Press.

Markham, P. L. (1985). The rational deletion cloze and global comprehension in German. *Language Testing*, *35*, 423–430.

Mokhtari, K., & Reichard, C. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, *94*, 249–259.

Mokhtari, K., & Sheorey, R. (2002). Measuring ESL students' awareness of reading strategies. *Journal of Development Education*, *25*, 2–10.

Nakatani, Y., & Goh, C. (2007). A review of oral communication strategies: focus on interactionist and psycholinguistic perspectives. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice* (pp. 207–227). Oxford, UK: Oxford University Press.

National College English Testing Committee. (2006). *CET–4 test syllabus and sample test paper* (2006 rev. version). Shanghai,China: Shanghai Foreign Language Education Press.

Oller, J. (1979). *Language tests at school: A pragmatic approach*. London, UK: Longman.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. New York, NY: Cambridge University Press.

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

Paris, S. G., & Jacobs, J. (1984). The benefits of informed instruction for children's reading awareness and comprehension skills. *Child Development*, *55*, 2083–2093

Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, *8*, 293–248.

Paris, S. G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15–51). Hillsdale, NJ: Erlbaum

Paris, S., Waskik, B., & Turner, J. (1991). The development of strategic readers. In R. Barr et al. (Eds.), *Handbook of reading research* (Vol. 2, pp. 609–640). New York, NY: Longman.

Pearson, P. D. (2009). The roots of reading comprehension instruction. In S. E. Isreal & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 3–31). New York, NY: Routledge.

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitve strategy use to EFL reading achievement test performance. *Language Testing*, *20*, 26–56.

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

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.

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.

Purpura, J. E. (1998). Investigating the effects of strategy use and second language test performance with high- and low-ability test takers: A structural equation modelling approach. *Language Testing*, *15*, 333–379.

Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modeling approach*. Cambridge, UK: Cambridge University Press.

Purpura, J. E. (2004). *Assessing grammar*. Cambridge, UK: Cambridge University Press.

Purpura, J. E. (2013). Cognition and language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1452–1476). Hoboken, NJ: Wiley/Blackwell.

Rumelhart, D. E. (2004). Toward an interactive model of reading. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 1149–1179). Newark, DE: International Reading Association.

Saito, Y. (2003). Investigating the construct validity of the cloze section in the Examination for the Certificate of Proficiency in English. *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, *1*, 39–82.

Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly*, *17*, 229–255.

Sheoery, R., & Mokhtari, K. (2001). Coping with academic materials: Differences in the reading strategies of native and non-native readers. *System: An International Journal of Educational Technology and Applied Linguistics*, *29*, 431–449.

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. (2011). Note-taking quality and performance on an L2 academic listening test. *Language Testing*, *29*, 67–89.

Song, X. (2005). Language learner strategy use and English proficiency on the Michigan English Language Assessment Battery. *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, *3*, 1–23.

Song, X., & Cheng, L. (2006). Language learner strategy use and test performance of Chinese learners of English. *Language Assessment Quarterly*, *3*, 243–266.

Stanovich, K. E. (1980). Towards an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*, 32–71.

Sticker, 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. Retrieved from http://www.ets.org/Media/Research/pdf/RR-05-12.pdf

Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled second language listener. *Language Learning*, *53*, 463–496.

Vandergrift, L., & Goh, C.M.C. (2012). *Teaching and learning second language listening: Metacognition in action*. New York, NY: Routlege.

Vandergrife, L., Goh, C. C. M., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The Metacognitive Awareness Listening Questionnaire: Development and validation. *Language Learning*, *56*, 431–462.

Wenden, A. L. (1998). Metacognitive knowledge and language learning. *Applied Linguistics*, *19*, 515–537.

Yang, H. Z., & Weir, C. J. (1998). *Validation study of the National College English Test*. Shanghai, China: Shanghai Foreign Language Education Press.

Zhang, L. J. (2010). A dynamic metacognitive systems account of Chinese university students' knowledge about EFL reading. *TESOL Quarterly*, *44*, 320–353.

Zhang, L. (2013). A structural equation modeling approach to investigating test takers' strategy use and their EFL reading test performance. *Asian EFL Journal*.

Zhang, L., Aryadoust,V., & Zhang, L.J. (in press) Development and validation of the Test Takers' Metacognitive Awareness Reading Questionnaire. *The Asia-Pacific Education Researcher*. doi: 10.1007/s40299-013-0083-z.

Zhang, L., & Zhang, L. J. (2013). Relationships between Chinese college test takers' strategy use and EFL reading test performance: A structural equation modelling approach. *RECL Journal*, *44*, 35–57.

Zheng, Y., & Cheng, L. (2008). Test review: College English Test (CET) in China. *Language Testing*, *25*, 408–417.

APPENDIX A

The Metacognitive and Cognitive Strategy Use Questionnaire

The purpose of this survey is to collect information about the various strategies you have used when taking the CET-4 Reading test you have just finished. Each statement is followed by five numbers, 0, 1, 2, 3, 4, and 5, and each number means the following:

- 0 means that "I **never** do this."
- 1 means that "I **almost never** do this."
- 2 means that "I do this **only occasionally**."
- 3 means that "I **sometimes** do this." (about **50%** of the time)
- 4 means that "I **usually** do this."
- 5 means that "I **always** or almost always do this."

After reading each statement, circle the number (0, 1, 2, 3, 4, or 5) which applies to you. Note that there are no right or wrong responses to any of the items on this survey.

When taking the CET-4 Reading subtest:

Strategy	*Never*					*Always*
1. I plan what to do before I start this reading test.	0	1	2	3	4	5
2. I make sure I am clear about the goals of the reading test task.	0	1	2	3	4	5
3. I think over essential steps needed to complete the reading test.	0	1	2	3	4	5
4. I read the title first and think over what the content of the text is about.	0	1	2	3	4	5
5. Test questions help me establish my purpose in reading.	0	1	2	3	4	5
6. I know what to do if my plan does not work well when I complete the reading test.	0	1	2	3	4	5
7. I critically evaluate the information presented in the text.	0	1	2	3	4	5
8. I evaluate my plan of test completion constantly.	0	1	2	3	4	5
9. I consider whether the content of the text fit my reading purpose.	0	1	2	3	4	5
10. I am aware of my loss of concentration in reading the text.	0	1	2	3	4	5
11. I infer what will happen next when reading the text.	0	1	2	3	4	5
12. I make summary of new information to understand the text better.	0	1	2	3	4	5
13. I take notes to increase my understanding.	0	1	2	3	4	5
14. I paraphrase (restate in my own words) to better understand the text better.	0	1	2	3	4	5
15. I know when I understand something and when I do not.	0	1	2	3	4	5
16. I know when I should complete the test more carefully.	0	1	2	3	4	5
17. I adjust my reading speed to increase comprehension.	0	1	2	3	4	5
18. I am aware when and where I am confused in the text.	0	1	2	3	4	5
19. I know when I should complete the test more quickly.	0	1	2	3	4	5
20. I budget my time wisely on this test.	0	1	2	3	4	5
21. I adjust pace in answering the questions.	0	1	2	3	4	5
22. I correct my misunderstanding or mistakes immediately when found.	0	1	2	3	4	5
23. I check my own performance and progress as I complete the test.	0	1	2	3	4	5
24. I use context clues to help me better understand the text.	0	1	2	3	4	5
25. I overview the text to see what it is about before reading it.	0	1	2	3	4	5

(*Continued*)

TABLE A1
(*Continued*)

Strategy	*Never*					*Always*
26. I preview the text first by noting its characteristics like length and organization.	0	1	2	3	4	5
27. I flip through the reading test before I actually start it.	0	1	2	3	4	5
28. I read the first sentence of each paragraph for the main idea.	0	1	2	3	4	5
29. I skip unknown words when reading.	0	1	2	3	4	5
29. I scan reading materials for specific words or phrases.	0	1	2	3	4	5
31. I use typographical features like boldface and italics to identify key information.	0	1	2	3	4	5
32. When the text becomes difficult, I reread the problematic part to increase my understanding.	0	1	2	3	4	5
33. If I understand some parts, I would use it as a clue to help me understand other parts.	0	1	2	3	4	5
34. I go back and forth in the text to find relationships among ideas in it.	0	1	2	3	4	5
35. I read the text not only for a surface understanding but also for its implied meaning.	0	1	2	3	4	5
36. I guess the meanings of new words from the context.	0	1	2	3	4	5
37. I make inference beyond the information presented in the text.	0	1	2	3	4	5
38. I try to use my prior knowledge to help my understanding.	0	1	2	3	4	5

APPENDIX B

Example Items of the College English Test Band 4 Reading Subtest

Part 1. Skimming and Scanning

The website for Orzack's center lists the following among the psychological symptoms of computer addiction:

- Having a sense of well-being or excitement while at the computer.
- Longing for more and more time at the computer.
- Neglect of family and friends.
- Feeling empty, depressed or irritable when not at the computer.
- Lying to employers and family about activities.
- Inability to stop the activity.
- Problems with school or job.

Physical symptoms listed include dry eyes, backaches, skipping meals, poor personal hygiene and sleep disturbances.

People who struggle with excessive Internet use may be depressed or have other mood disorders, Orzack said. When she discusses Internet habits with her patients, they often report that being online offers a "sense of belonging, an escape, excitement [and] fun," she said. "Some people say relief . . . because they find themselves so relaxed."

1. According to Orzack, people who struggle with heavy reliance on the Internet may feel_____.

A. depressed B. pressured C. discouraged D. puzzled

Part 2. Banked Cloze

Fortunately, there are a __1__ number of relatively simple changes that can green older homes, from __2__ ones like Lincoln's Cottage to your own postwar home. And efficiently upgrades can save more than just the earth; they can help __3__ property owners from rising power costs.

A) accommodations	I) protect
B) clumsy	J) reduced
C) doubtfully	K) replace
D) exceptions	L) sense
E) expand	M) shifted
F) historic	N) supplying
G) incredibly	O) vast
H) powering	

Part 3. Reading in Depth

You never see them, but they're with you every time you fly. They record where you're going, how fast you're traveling and whether everything on your airplane is functioning normally. Their ability to withstand almost any disaster makes them seem like something out of a comic book. They are known as the black box.

1. What does the author say about the black box?

A) Its ability to ward off disasters is incredible.
B) It is an indispensible device on an airplane.
C) It ensures the normal functioning of an airplane.
D) The idea for its design comes from a comic book.

Part 4. Multiple-Choice Cloze

The terms e-commerce refers to all commercial transactions conducted over the Internet, including transactions by consumers and business-to-business transactions. Conceptually, e-commerce does not __1__ from well-known commercial offerings such as banking by phone, "mail order" catalogs, or sending a purchase order to a supplier__2__ fax. E-commerce follows the same model __3__ in other business transactions; the difference __4__ in the details.

1 A) distract	B) differ	C) derive	D) descend
2 A) off	B) from	C) via	D) with
3 A) appeared	B) used	C) resorted	D) served
4 A) roots	B) lies	C) locates	D) situates

Note. Due to limited space, only example test items are listed in Appendix B.

APPENDIX C

Standardized Direct Effects and Error/Disturbance of the Unitary Model

	Direct Effects										
Variable	*STR_U*	*LEX_GR*	*TxtCOM*	*MON*	*EVA*	*PLA*	*INF*	*INT*	*IDE*	*INI*	*Error/ Disturbance*
LEX_GR	.16*										.97*
TxtCOM	.04	.88*									.21*
MON	.76*										.43*
EVA	.68*										.54*
PLA	.51*										.74*
INF	.63*										.60*
INT	.60*										.64*
IDE	.64*										.59*
INI_REA	.66*										.57*
SKSN			.61*								.63*
RID			.65*								.58*
BCLZ		.70*									.51*
MCLZ		.82*									.33*

*$p < .05$.