Language Testing 2008 25 (2) 237–272

Construct validation of Bachman and Palmer's (1996) strategic competence model over time in EFL reading tests

Aek Phakiti *University of Sydney, Australia*

This article reports on a large-scale study that aims to validate the theory of strategic competence proposed by Bachman and Palmer (1996) through the use of structural equation modeling (SEM). The present study examines the relationship of test-takers' long-term strategic knowledge (i.e., trait strategies) and actual strategy use (i.e., state strategies) to second language (L2) reading test performance over time. The data were gathered on two occasions (during the mid-term and final examination periods). Five hundred and sixty-one Thai university test-takers answered a trait strategy use questionnaire prior to the mid-term and final reading achievement tests and, immediately after completing each test, they answered a state strategy use questionnaire. It was found that (1) trait metacognitive strategy use (MSU) directly and strongly affects trait cognitive strategy use (CSU); (2) trait CSU does not greatly affect state CSU; (3) trait MSU directly affects state MSU in a specific context, which in turn directly affects state CSU; and (4) state CSU directly affects a specific language test performance to a varying degree.

Key words: second language reading performance, state and trait strategy use, strategic competence, structural equation modeling

I Introduction

McNamara (1996) pointed out three basic dimensions in which the nature of second language (L2) communicative ability has been conceptualized: (1) factors constituting knowledge of a language; (2) other non-linguistic factors that are not specific to language such as strategic competence and affect; and (3) the way actual real-time instances of language use are seen in the light of the preceding dimensions. Inferences about the first and second dimensions, according to McNamara, rest on an understanding of how language

Address for correspondence: Aek Phakiti, TESOL Program, Faculty of Education and Social Work, University of Sydney, NSW 2006, Australia; email: a.phakiti@edfac.usyd.edu.au

 DOI:10.1177/0265532207086783

knowledge interacts with the other non-linguistic factors constituting *ability for use* (Hymes, 1972). Bachman and Palmer's (1996) work is evidence of an attempt to develop a model of what is involved in such dimensions. In Bachman and Palmer's (1996) model, communicative language ability (CLA) is hypothesized to be the major contributor to language performance (other contributors include test-method facets, individual characteristics and random factors). In their model, Bachman and Palmer (1996) separate strategic competence (as a non-linguistic factor) from language competence which, according to McNamara (1996), is an important step in discussing L2 communicative ability. Strategic competence is a general ability that enables an individual to use available resources by regulating online cognitive processes in accomplishing a communicative goal (e.g., assessing the situation, setting goals and planning what to do).

Despite a decade of research into Bachman and Palmer's powerful theoretical model, much remains unclear in regard to the exact nature of strategic competence. McNamara (1996) pointed out that discussion of strategic competence in this model is extremely preliminary and recommended that a full discussion of the model would be helpful if cross-referenced to the literature in the areas of psychology, social psychology and pragmatics. In addition, Purpura (1999) pointed out that the depiction of strategic competence in Bachman and Palmer's (1996) model is not based on empirical research. Therefore, Bachman and Palmer's (1996) strategic competence model needs validation with empirical data. To date, no one has attempted to validate Bachman and Palmer's (1996) strategic competence model comprehensively with empirical longitudinal data from high-stakes test contexts.

II Review of the literature

1 Validation research on the strategic competence model

There is consensus that strategic processing has a component of *awareness* and occurs within the working memory realm (stipulated within the *focal attention* or at least within *peripheral attention*; Cohen, 2007; Schmidt, 2001). What may distinguish strategies from skills and other processes is the level of *awareness and deliberation on the processes*, rather than the nature of processes per se (Alexander *et al.*, 1998). In L2 reading, for example, some individuals may paraphrase, underline, highlight or summarize the text without realizing they do, whereas others may be conscious and purposeful in using such processes. Despite the fact that strategy research ranks among the more popular topics in applied linguistic research, drawing conclusions about the nature of strategic processing from language learning strategy research (see Anderson, 2005) in light of Bachman and Palmer's (1996) strategic competence model is not always straightforward. First, there are many schools of thought on strategic processing which may not attempt to produce or fine-tune the theoretical concept of strategic competence in L2 communicative ability (e.g., O'Malley & Chamot, 1990; Oxford, 1990). A review of language learning strategy research shows that little effort has been made to distinguish observable strategic behaviors (e.g., cognitive and metacognitive strategy use) from the unobservable strategic competence construct that forms part of L2 communicative ability. Furthermore, test-taking strategy research (see Cohen, 1998, 2007) that investigates special techniques test-takers employ to optimize their language test performance has been oriented towards test development, rather than used to validate strategic competence theory. Cohen (2007) further points out that strategy data, particularly in validation research, are not usually collected in actual high-stakes testing situations. It may be that strategies actually used in responding to tests in high-stakes situations differ from those identified under research conditions (Cohen, 2007), because there is no consequence for not answering test items.

In the language testing (LT) literature, there are only a few published studies that aim to validate Bachman and Palmer's (1996) strategic model. Grounding his study in human-information processing theory (Gagné, Yekovich & Yekovich, 1993) which facilitates generalizations about strategic processing in L2 use, Purpura (1999) examined the relationship between perceived cognitive and metacognitive strategy use and language test performance (UCLES's First Certificate in English (FCE) Anchor Test). He explored this relationship through the application of the structural equation modeling (SEM) approach with 1382 learners who answered a context-free strategy use questionnaire prior to test-taking. It was found that cognitive processing was a *multidimensional* construct consisting of a set of *comprehending*, *memory* and *retrieval* strategies. These complex cognitive strategies worked with one another to affect language performance. The SEM model of metacognitive strategy use was found to be a *unidimensional* construct consisting of a single set of assessment processes (e.g., goal setting, planning, monitoring, self-evaluating and self-testing). Purpura found that metacognitive processing had significant, direct and positive effects on all the three components of cognitive processing (values between 0.59 and 0.86) which directly affected language test performance. This finding suggests that the effect of cognitive strategies on test performance is mediated by metacognitive strategies. Purpura (1999) also found that the high-ability test-takers were likely to use metacognitive strategies more automatically than the low-ability ones. These different patterns in turn had a significant impact on their language performance. In his conclusion, Purpura (1999) pointed out that Bachman and Palmer's (1996) strategic competence notion must be extended beyond a set of metacognitive strategies because individuals invoke cognitive, affective and social strategies as well as metacognitive strategies when they use the target language.

Following the work of Purpura (1999), Phakiti (2003b), through the use of a cognitive and metacognitive questionnaire drawn from the existing literature, retrospective interviews and an English-as-a-foreign-language (EFL) achievement test, investigated the relationship between 384 Thai learners' cognitive and metacognitive strategy use and their reading test performance. His study differed from Purpura's in that test-takers answered the questionnaire regarding the degree of their strategy use during the test-taking immediately after test completion. Using the factor structures from exploratory factor analyses (EFAs) to form composites of cognitive and metacognitive strategies for further quantitative analyses, Phakiti found that metacognitive strategies were statistically positively related to cognitive strategies (at a correction-for-attenuation correlation of 0.76). With regard to the relationships between strategies and test performance, cognitive and metacognitive strategy use was positively correlated with the reading test performance, explaining about 15–22% of the test score variance. Phakiti also found statistical differences in the reported use of cognitive and metacognitive strategies between highly successful, moderately successful and unsuccessful test-takers.

Other studies that examine the extent to which cognitive and metacognitive strategy use is related to L2 test performance include those of Song (2004 and 2005). Song (2004), by means of a revised strategy questionnaire mainly based on Purpura (1999), investigated the extent to which cognitive and metacognitive strategy use accounted for Chinese test-takers' performance in the College English Test Band 4 through regression analyses. Song found that cognitive and metacognitive strategies accounted for 8.6% of the test score. In the context of the Michigan English Language Assessment Battery (MELAB), Song (2005), after examining the strategy structures by means of EFAs and through regression analyses, found that the effects of strategy use on language performance were weak to moderate (explaining about 12.5–21.4% of the score variance).

2 Implications from the current research

Findings and insights from previous LT research have at least four major implications for the design of the present study. The first implication is that previous research may have examined different facets of strategic competence. Purpura (1999) and Song (2005) examined test-takers' perceptions regarding their normal use of a set of cognitive and metacognitive strategies without reference to a specific context, while Phakiti (2003b) examined test takers' reported cognitive and metacognitive strategy use in a specific test context. The strategy use questionnaire items in Purpura, on the one hand, are written using the *Simple Present* tense, which reflects individuals' habitual strategy use, for example, 'I *double-check* my understanding when I *read*.' Each strategy use item in Phakiti, on the other hand, is written using the *Simple Past* tense, which suggests that the use of the strategy relates specifically and exclusively to a particular context and occasion, for example, 'I *double-checked* my understanding during this reading test.' Accordingly, there may be underlying assumptions about the nature of strategic competence that need clarification prior to empirical data gathering.

Although Bachman and Palmer (1996) do not refer to strategic competence as metacognition, it can be argued that the nature of the two constructs is similar, given that both concern self-regulation and hence, as McNamara (1996) pointed out, the strategic competence model should be cross-referenced with the literature in the areas of metacognition. According to metacognitive research (Baker & Brown, 1984; Flavell, 1985), metacognition has dual components: *knowledge about cognition*, which is general information that a person possesses concerning his or her awareness about strategy use when engaged in activities in general, and *regulation of cognition*, which refers to the effectiveness with which the person keeps track of ongoing cognitive processes and his or her use of strategies in order to solve problems or achieve desired goals. There is consensus in metacognitive research that these two components are separate phenomena. That is, knowledge about cognition is relatively stable and stored in long-term memory (LTM), whereas regulation of cognition, such as checking comprehension and evaluating performance, is rather unstable due to the nature of specific tasks and contexts at hand and occurs within working memory (WM) space. Baker and Brown (1984) pointed out that the former does not essentially equate to or fully guarantee the latter. It may be argued that both individuals' knowledge about cognition and their capacity for regulation of cognition are what constitute strategic competence.

According to anxiety research (Spielberger, 1972), state and trait notions may be applied in the assessment of the two metacognitive constructs (knowledge of cognition and regulation of cognition). In psychology, states and traits refer to two different classes of individuals' psychological attributes which include (1) a relatively stable trait and (2) a transitory state. A trait facet of a construct is a relatively stable attribute of an individual across occasions (despite considerable variation in the range of settings and circumstances), whereas a state facet is transitory, fluctuating and unstable in a given context. Hence, knowledge about cognition represents trait metacognition, whereas regulation of cognition represents state metacognition. In L2 strategy research, general perceived strategy use free of contexts is trait-like knowledge of cognition, whereas actual perceived strategy use in a specific context is state-like regulation of cognition. Though not qualifying as traits and states in the original sense of the terms in anxiety research (Spielberger, 1972), the use of state and trait notions to classify the two aspects of strategy use in strategic competence research is a way forward. As Phakiti (2003a) has proposed, an empirical research into the relationship between perceived strategy use across contexts and actual strategy use in a particular context has the potential to offer greater insights into an individual's psychology.

For the purpose of this study, cognitive strategies are defined as actual conscious behaviors that individuals use to process language to understand, learn or use in some context. Following the classifications of cognitive and metacognitive strategy use[1] by Purpura (1999), which are based on the theory of human-information processing (Gagné, Yekovich & Yekovich, 1993), cognitive strategies are composed of *comprehending* (for understanding, such as identifying main ideas, author's attitudes, translation, predicting, inferencing),

[1]It should be noted that Bachman and Palmer (1996) do not discuss cognitive strategies in their model. Rather, they discuss strategic competence as metacognitive strategies which have executive functions in language use. However, as Purpura (1999) pointed out, in order to better understand strategic competence, cognitive and metacognitive strategy use should be included. The present study, thus, does not validate Bachman and Palmer's (1996) original strategic competence model, but rather it validates the more finely tuned conceptualization of their original theory that has evolved over the past decade.

memory (for storing information in memory, such as rereading or repeating, note taking or underlining, paraphrasing) and *retrieval* (for recalling information, such as using prior knowledge/experience, applying grammatical rules/knowledge) strategies. Metacognitive strategies are conscious processes that regulate cognitive strategies and other processing. They are composed of *planning* (for future actions and goal attainment, such as goal-setting, overseeing tasks, planning actions beforehand), *monitoring* (for checking ongoing comprehension or performance, such as noticing comprehension failure or errors, double-checking comprehension) and *evaluating* (for evaluation of past and current actions or performance, such as assessing level of difficulty, self-questioning, evaluating performance/product accuracy) strategies. Hence, trait cognitive and metacognitive strategy use is a general tendency of an individual to use cognitive and metacognitive strategies over a variety of contexts, whereas state cognitive and metacognitive strategy use is a specific intensity of cognitive and metacognitive strategy use in an actual language use situation.

The second implication is that previous research was cross-sectional and might have limited generalizability, if we accept that cognitive processing unfolds over time and depends greatly on the context in which it occurs. Since cross-sectional designs allow only for the assessment of relationships among variables at one point in time, they do not allow for autoregressive effects or time lags (MacCallum & Austin, 2000), thereby limiting inferences on causality or directional strategy influence on language learning or use. Furthermore, L2 test performance has been acknowledged as highly complex, multidimensional and variable according to a variety of social and contextual factors (McNamara, 1996). This complexity cannot be captured in a single performance (Spolsky, 1995). Hence, there is a need to sample it across a range of contexts in order to demonstrate both behavioral consistency and factors involved at a given time. Observation of language performance and strategy use consistency is fundamental to the study of L2 test performance (Chapelle, 1998). Chapelle (1998) argued that operational settings change according to changes in test method facets which in turn influence individuals' construction of context and their use of metacognitive strategies. A well-known research tool to investigate the effect of context change is the multitrait-multimethod (MTMM) approach (Campbell & Fiske, 1959) which is adopted in the present study. In this study, however, only two data points are obtained due to the fact that there were only two high-stakes tests (mid-term and final tests) in this context (which is a limitation of the present study).

The third implication is that some previous research employed standard statistical approaches (e.g., correlational and regression analyses) which are subject to certain analytical limitations since they do not directly explore and evaluate error of measurement effects which limit generalization. Hence, following the work by Purpura (1999), the SEM approach – a robust statistical analysis for testing substantive theories – will be employed because it directly considers score reliability (i.e., [1 – measurement error variance] / total score variance) as part of model fitting. Since all measures (tests and questionnaires) contain non-random error, it is important to allow this error to be incorporated in the models in such a way that will not directly affect parameter estimates (Purpura, 1999). The SEM approach is powerful in that it analyzes the structure and effect of unobservable latent variables through the analysis of individual differences data, by statistically relating co-variation between observed variables to latent variables. The underlying principle of this analysis is that the statistical fit of such a model can be adduced as evidence supporting the hypothesis.

The fourth implication is that because various aspects of language are processed and stored differently in the brain for different language skills (VanPatten, 1994), different domains of language need to be specifically researched (Schmidt, 1995). Hence, EFL reading comprehension tests are chosen. Another reason for excluding other language skills is due to the depth and maturity of L2 reading research which allows the researcher to go beyond exploratory, baseline investigations. A number of studies in L2 reading strategies have provided insights into how strategic reading affects success in L2 comprehension and how good readers differ from poor readers due to the strategic ability variable (see Alderson, 2000; Anderson, 1999; Carrell *et al.*, 1998; Hudson, 2007).

3 Research questions and hypotheses

The literature above has suggested that first, metacognitive strategy use (MSU) is directly and highly related to cognitive strategy use (CSU). MSU such as planning, monitoring and evaluating regulates the extent to which cognitive strategies are to tackle specific needs in language use. Second, cognitive strategies may have a direct relationship to a particular language performance since they are more closely related to the target language than metacognitive strategies. The SEM models tested by Purpura are extended to include both state and trait strategy use in this study and to accommodate over time data collections (two months apart) for observation of consistency in findings.

State CSU is hypothesized to directly influence a specific test performance. With regard to language test performance, lexico-grammatical ability as assessed through gap-filling reading tests is hypothesized to influence reading comprehension ability (e.g., text comprehension), since success in L2 reading largely depends on decoding and the activation of lexico-grammatical knowledge (Nation, 2001) and learners at upper-beginners and intermediate levels need to decode texts at the lexico-grammatical levels when they read (Purpura, 1999). Figure 1 presents the hypothesized model. The model in Figure 1 is derived from a combination of two separate SEM analyses (i.e., Time 1 and Time 2; to be discussed in the Methods section).[2] The longitudinal SEM modeling (as in Figure 1) has at least three main advantages for this validation research: it allows us (1) to examine the co-variation of variables over time, (2) to test models that include data collected at multiple time points, thereby making it possible to test bi-directionality of cause and effect, and (3) to examine the consistency of variables' effects on others over time. The hypothesized model will answer the following research questions:

- To what extent does trait MSU at Time 1 relate to trait MSU at Time 2?
- To what extent does trait MSU affect trait CSU?
- To what extent does trait CSU affect state CSU?
- To what extent does trait MSU affect state MSU in a specific context?
- To what extent does state MSU affect state CSU?
- To what extent does state CSU affect a specific language test performance?

III Methods

1 Setting and participants

The study was conducted at a Thai government university in the north of Thailand. Bachelor's degree students enrolled in an English

---

[2]During model modification, the Lagrange Multiplier (LM) Test (to test the hypothesis on statistical necessity of restrictions existing in the hypothesized model, i.e., for adding parameters) suggested a missing path between trait MSU at Time 1 and trait MSU at Time 2. Given that trait MSU should be related, a correlation path was added in the model. Furthermore, the direct influence of trait MSU and trait CSU on language test performance was tested prior to the model in Figure 1. However, the model did not hold statistically; the Wald (W) test for dropping parameters suggested that the parameters were non-significant to the model; and hence the model was rejected.

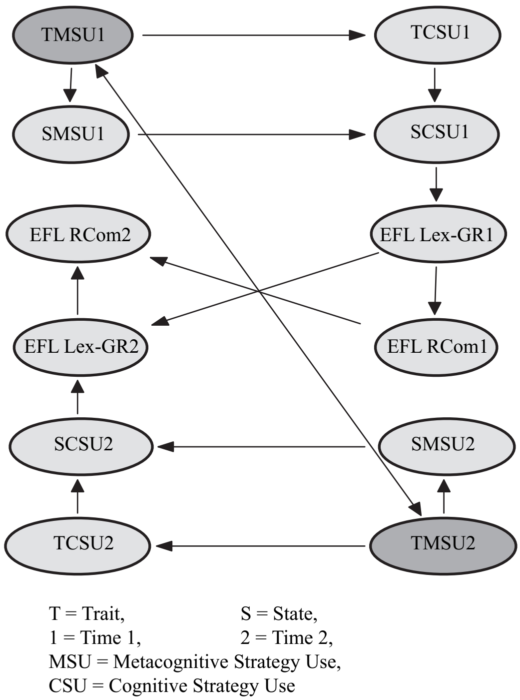


Figure 1 Hypothesized relationships between state and trait cognitive and metacognitive strategy use and L2 reading test performance over time

subject (*Fundamental English I*) were asked to voluntarily participate in the study. The data were gathered during their mid-term (accounting for 35% of the total grade) and final (40%) test periods. Data from 561 Thai university students were used in this study. They were made up of 201 males (37%) and 360 females (63%). They were between the ages of 18 and 24 (mean age = 18.53, SD = 0.70). All of these test-takers had been studying English in Thailand for approximately eight years. Based on their English entrance examination results, their English proficiency levels ranged from advanced beginners to intermediate.

2 Measurement instruments

There were two major sets of research instruments in this study: (1) two different English reading tests (mid-term and final tests); and (2) four 6-point Likert-scale, strategy use questionnaires (2 states, 2 traits) in Thai.

a Mid-term and final reading test: The topics in the tests include family, occupation, clothing and fashion, personalities, giving directions, accommodation, food and drink, travel and transportation. The mid-term and final tests were multiple-choice, 80-item tests. Each test lasted three hours. The mid-term and final reading tests were composed of two major sections with sub-sections in them.

- *Section 1: Gap-filling* (like rational cloze): This section (composed of three sub-sections) was designed to measure the test-takers' ability to comprehend texts drawing on their knowledge of *structural* and *lexical* appropriacy, as well as their pragmatic and discourse competence (see Read, 2000). Items tested were selected based on the structural, lexical, pragmatic and discourse skills taught in the class. In the mid-term test, there were 46 test items (21 to measure reading/vocabulary; 25 to measure reading/grammar). In the final test, there were 43 test items (20 to measure reading/vocabulary; 23 to measure reading/grammar). For the purpose of this study, performance on this section was labelled as lexico-grammatical reading ability (LEX-GR).
- *Section 2: Reading comprehension*: This section (composed of 2 sub-sections) consisted of various passages, ranging from 100 words to 700 words. It aimed to measure the test-takers' ability to read English texts for main ideas, details and inferences. The first sub-section aimed to measure scanning and skimming for information abilities. The second sub-section aimed to measure the ability to identify main topics/ideas, titles, writer's purposes, reference words, implied statements, vocabulary in context, true/false statements and specific details. In the mid-term test, there were 34 items for Section 2 and in the final test, there were 37 items. For the purpose of this study, performance on this section was labelled as text comprehension (TxtCOM).

Approximate person-separation estimates (like KR-20; based on the Rasch IRT analyses by the Quest Program; Adams and Khoo, 1996) were 0.91 for the mid-term test (0.86 and 0.83 for Sections 1 and 2, respectively) and 0.89 for the final test (0.82 and 0.79 for Sections 1 and 2, respectively). Based on misfit statistics derived from the IRT analyses, misfitting test-takers in both the mid-term and final tests were eliminated from the data set. Table 1 presents descriptive statistics of the 10 observed reading test performance variables to be used in the SEM models. All variable skewness and kurtosis statistics were within the acceptable limits, which was suggestive of univariately normal distributions. The mid-term test variables (LEX-GR1,

Table 1 Descriptive statistics of the five observed reading test performance variables based on the mid-term and final tests

Item	No. of items	Min	Max	Mean	SD	Median	Mode	Skewness	Kurtosis
LEX-GR1	10	0.00	10.00	5.92	2.15	6.00	5.00	−0.05	−0.58
LEX-GR2	13	2.00	13.00	8.31	2.39	8.00	9.00	−0.04	−0.68
LEX-GR3	23	4.00	23.00	14.80	4.07	14.00	13.00	−0.00	−0.72
TxtCOM1	14	4.00	14.00	10.35	2.16	10.00	10.00	−0.30	−0.37
TxtCOM2	20	2.00	20.00	10.66	4.38	10.00	8.00	0.53	−0.68
LEX-GR4	9	1.00	9.00	5.39	1.56	5.00	6.00	0.05	−0.22
LEX-GR5	16	0.00	16.00	8.12	2.96	8.00	8.00	0.51	−0.07
LEX-GR6	18	1.00	18.00	8.55	3.30	8.00	7.00	0.59	0.12
TxtCOM3	20	2.00	20.00	11.58	3.30	12.00	11.00	−0.05	−0.53
TxtCOM4	17	1.00	17.00	8.55	3.14	8.00	8.00	0.20	−0.47

LEX-GR = Lexical-Grammatical ability; **TxtCOM** = Text comprehension ability;
LEX-GR1, LEX GR2, LEX GR, TxtCOM1 and TxtCOM2 = mid-term test variables (Time 1);
LEX-GR4, LEX-GR5, LEX-GR6, TxtCOM3 and **TxtCOM4** = final test variables (Time 2)

LEX-GR2, LEX-GR3, TxtCOM1 and TxtCOM2) represent Time 1 and the final test variables (LEX-GR4, LEX-GR5, LEX-GR6, TxtCOM3 and TxtCOM4) represent Time 2.

b State and trait reading strategy questionnaires: Strategy use items were selected from the literature (e.g., Cohen, 1998; Hsiao & Oxford, 2002; Mokhatari & Sheorey, 2002; O'Malley & Chamot, 1990; Oxford, 1990; Phakiti, 2003b; Purpura, 1999). The items were chosen based on the theory of human information processing (Gagné *et al*., 1993) which postulates (1) a structural component of sensory receptors, working and long-term memory arrays and (2) a functional component of information processing that describes the operations of comprehending, memory, retrieval and control processes at different specific stages. The state and trait strategy questionnaires were developed and validated prior to actual use (i.e., by means of trials with a similar target population, reliability and exploratory and confirmatory factor analyses, and expert judgments). It was found that the components of the questionnaire aligned with empirical factor analysis findings. Table 2 presents the strategy composites in the questionnaire.[3]

[3] The original questionnaires included five affective strategies (Items 31–35) which were excluded from the analyses reported in the present article. The reasons for excluding these items were mainly because the number of observed variables was not large enough to statistically form a latent affective strategy use composite.

Table 2 Taxonomy of cognitive and metacognitive strategies

Processing	Sub-scale		No. of items	Items
1. Cognitive strategies	Comprehending		4	7, 8, 9, 10,
	Memory		4	11, 12, 13, 15
	Retrieval		5	14, 16, 17, 18, 19
		Subtotal	13	
2. Metacognitive strategies	Planning		6	1, 2, 3, 4, 5, 6
	Monitoring		5	20, 21, 24, 29, 30
	Evaluating		6	22, 23, 25, 26, 27, 28
		Subtotal	17	
		Total	30	

The questionnaires allowed the participants to mark their strategy use on a 6-point Likert scale: 0 (Never), 1 (Rarely), 2 (Sometimes), 3 (Often), 4 (Usually) and 5 (Always). The questionnaires were then translated into Thai, so that the contents were comprehensible for all participants. Appendix A provides the English version of the state and trait strategy questionnaires. Trait strategy use items are written using the Simple Present, whereas state strategy use items are written using the Simple Past.

For all six strategy data sets, the items belonging to each strategy composite were tested by confirmatory factor analyses (CFAs), and the statistics supported the structure of each underlying factor of the strategies in this study. A composite of each strategy category (as in Table 2) was generated by combining scores from the designated strategy items and dividing the total score by the number of strategy items in the relevant set. For example, scores for Items 7, 8, 9 and 10 were combined and divided by four to form the comprehending strategy composite. This generation method can minimize random response and transient errors (Schmidt & Hunter, 1999). Table 2 presents the summary descriptive statistics for the 24-strategy composites. All values for variable skewness and kurtosis were within the limits (all values near zero) which suggested univariately normal distributions. Tables 3 and 4 present the composites of the eight state and trait latent variables with internal consistency estimates (Cronbach alpha). The composite of each latent variable was derived from Table 2.

3 Data collection

Figure 2 presents a flow chart of the data gathering procedures. The participants answered a trait strategy use questionnaire about *one*

Table 3 Descriptive statistics for the state and trait cognitive, metacognitive and affective strategy variables

Item	Min.	Max.	Mean	SD	Median	Mode	Skewness	Kurtosis
TPLN1	1.33	5.00	3.24	0.83	3.17	2.67	0.26	−0.74
TMON1	1.80	5.00	3.71	0.71	3.60	3.60	−0.07	−0.59
TEVA1	1.33	5.00	3.32	0.79	3.33	3.00	0.13	−0.56
TCOM1	1.00	5.00	3.27	0.82	3.25	3.00	−0.01	−0.63
TMEM1	1.00	5.00	3.22	0.81	3.25	2.75	0.08	−0.53
TRET1	0.40	5.00	3.13	0.81	3.00	2.80	0.06	−0.31
SPLN1	1.17	5.00	3.25	0.71	3.33	3.50	−0.09	−0.41
SMON1	1.60	5.00	3.69	0.70	3.80	4.00	−0.24	−0.53
SEVA1	1.33	5.00	3.38	0.72	3.33	3.67	−0.00	−0.49
SCOM1	1.25	5.00	3.34	0.72	3.25	3.25	0.01	−0.45
SMEM1	0.75	5.00	3.25	0.73	3.25	3.50	−0.05	−0.38
SRET1	1.20	5.00	3.20	0.73	3.20	3.20	0.00	−0.18
TPLN2	1.50	5.00	3.33	0.70	3.33	3.67	0.19	−0.39
TMON2	1.80	5.00	3.69	0.73	3.80	4.00	−0.20	−0.56
TEVA2	1.33	5.00	3.37	0.68	3.33	3.33	0.06	−0.03
TCOM2	1.00	5.00	3.33	0.72	3.25	3.00	0.08	−0.26
TMEM2	1.50	5.00	3.38	0.70	3.38	3.50	−0.14	−0.25
TRET2	1.20	5.00	3.22	0.71	3.20	3.60	0.03	−0.23
SPLN2	1.67	5.00	3.34	0.65	3.33	3.33	0.14	−0.33
SMON2	1.80	5.00	3.67	0.70	3.60	4.00	−0.09	−0.54
SEVA2	1.83	5.00	3.45	0.65	3.50	3.33	0.04	−0.48
SCOM2	1.50	5.00	3.36	0.65	3.25	3.25	0.00	−0.39
SMEM2	1.50	5.00	3.41	0.63	3.50	3.25	−0.09	−0.42
SRET2	1.40	5.00	3.29	0.65	3.20	3.40	0.07	−0.27

T = trait; **S** = state; **PLN** = planning; **MON** = monitoring; **EVA** = evaluating; **COM** = comprehending; **MEM** = memory; **RET** = retrieval; **1** = Time 1; **2** = Time 2

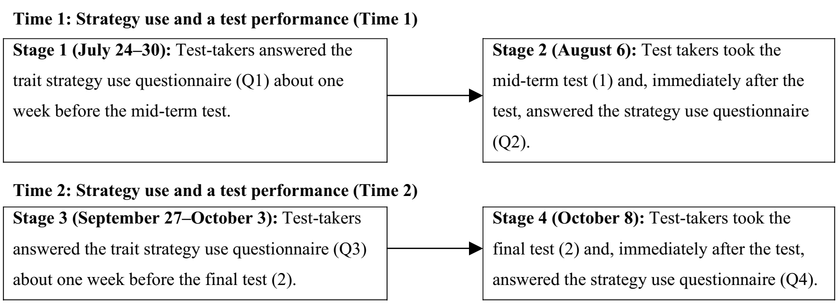


Figure 2 A flow chart of data-gathering procedures

week before each test was given. The length of time to complete the questionnaire was approximately 15 to 25 minutes. State strategies were immediately measured after learners had completed the test. The final test occurred about two months after the mid-term test.

Table 4 Internal consistency estimates (Cronbach alpha) for observed variables

Composite	No. of items	Items used	Internal consistency
1. Trait metacognitive strategy use Time 1	3	TPLN1, TMON1, TEVA1	0.86
2. Trait cognitive strategy use Time 1	3	TCOM1, TMEM1, TRET1	0.89
3. State metacognitive strategy use Time 2	3	SPLN1, SMON1, SEVA1	0.83
4. State cognitive strategy use Time 2	3	SCOM1, SMEM1, SRET1	0.88
5. Lexical-Grammatical ability Time 1	3	LEX-GR1, LEX-GR2, LEX-GR3	0.86
6. Reading comprehension ability Time 1	2	TxtCOM1, TxtCOM2	0.83
7. Trait metacognitive strategy use Time 2	3	TPLN2, TMON2, TEVA2	0.87
8. Trait cognitive strategy use Time 2	3	TCOM2, TMEM2, TRET2	0.89
9. State metacognitive strategy use Time 2	3	SPLN2, SMON2, SEVA2	0.86
10. State cognitive strategy use Time 2	3	SCOM2, SMEM2, SRET2	0.87
11. Lexical-Grammatical ability Time 2	3	LEX-GR4, LEX-GR5, LEX-GR6	0.82
12. Reading comprehension ability Time 2	2	TxtCOM3, TxtCOM4	0.79
Total	34		0.95

4 SEM analyses

SPSS (Statistical Packages for Social Sciences) program was used as a data file manager prior to using the EQS 6 program for SEM (Bentler, 1985–2006). The EQS 6 program was used to impute data, perform missing and outlier data analyses, confirmatory factor analyses (CFAs), covariance structure analyses and finally SEM. The maximum likelihood (ML) estimation method was employed in model estimations. ML estimation is typically used to seek parameters that best reproduce the estimated population variance-covariance matrix. The evaluation of model adequacy is based on an inspection of the values of standardized residuals, the chi-square (χ^2) statistics, other fit indices (e.g., Bentler-Bonett normed fit index (NFI) and comparative fit index (CFI) and informed by my knowledge of the data and theoretical and conceptual aspects of the constructs under study (see Bentler, 2006; Byrne, 1994, for a discussion of the evaluation of model adequacy). It was found that the changes in the factor loadings on both occasions in the full SEM model (Figure 3) in comparison with those in each separate cross-sectional SEM model (see Appendix B) were statistically non-significant. Based on Purpura (1999), changes in factor loadings and parameter estimates in a full model are to be expected as more information is included for analysis.

IV Results and discussion

1 Statistical evaluation of the hypothesized SEM model

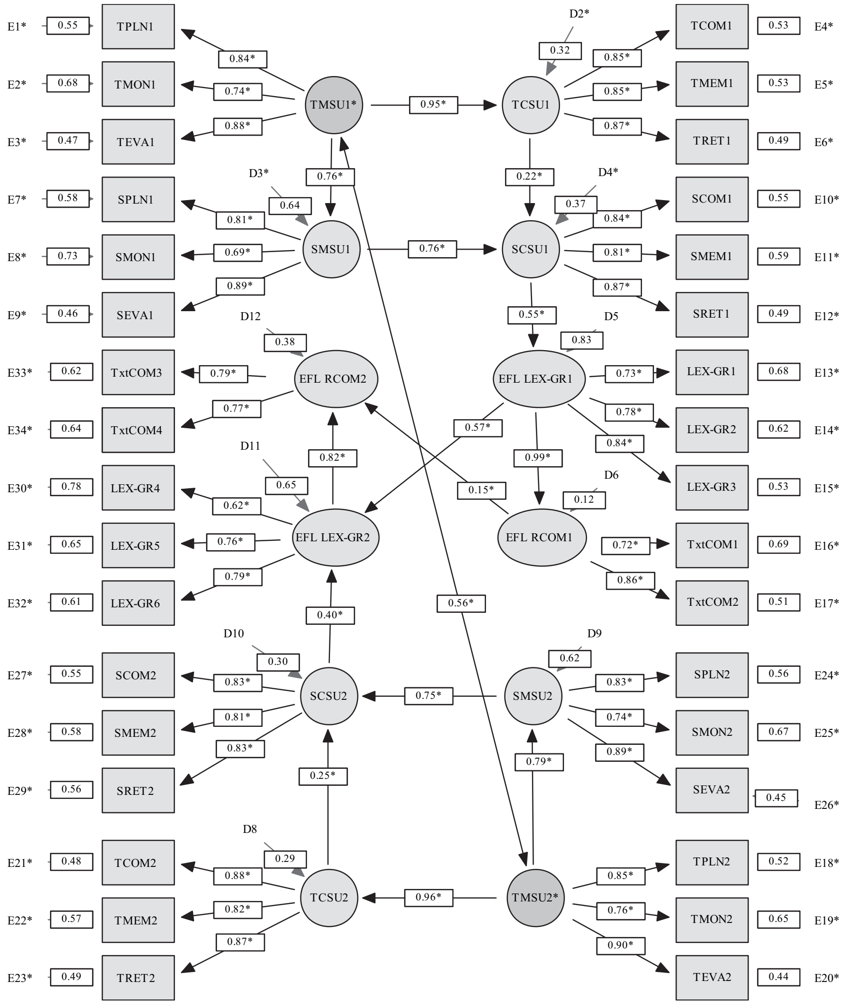
It was found that the global fit of the model was adequate for interpretation of the relationships between state and trait strategy use and EFL reading test performance over time. For example, the average off-diagonal absolute standardized residual was 0.063 which was near zero. The independence chi-square statistic ($\chi^2_{(561)}$) was 15583.483. The Chi-square statistic of the full SEM model ($\chi^2_{(500)}$) was 1802.437 (as in Figure 3). The probability value for this chi-square statistic was significant ($p < 0.000$).[4] In the SEM approach, the large difference between the independence χ^2 statistic and that of the full model (i.e., 13781.046) suggested a good model fit (Byrne, 1994). The standard fit indices such as the Bentler-Bonett Normed Fit Index (NFI) which was 0.90, the Bentler-Bonett Non-Normed Fit Index (NNFI) which was 0.91, the Comparative Fit Index (CFI) which was 0.93 and Root Mean-Square Error of Approximation (RMSEA) which was 0.058 (ideally RMSEA should be less than 0.05), were acceptable. Generally speaking, fit indices of 0.90 and above are deemed sufficient to accept the tested model (Bentler, 2006) as such values indicate that the relative overall fit of the hypothesized model is about 90% better than that of the null model estimated with the same sample data. The values of the individual parameters were inspected and it was found that all were within the expected ranges (± 1 range, $\pm$ SE for standardized solution) and there were no anomalies.

2 Parameter estimation

Figure 3 presents the hypothesized SEM model along with the estimates of factor loadings and error/disturbance terms. Estimates

[4]Unlike other standard statistical analyses, SEM researchers need to obtain a nonsignificant χ^2 ($p > 0.001$). Limitations of the χ^2 statistic as a fit index (e.g., no upper bound, sensitivity to sample size) are widely discussed in most SEM textbooks (e.g., Byrne, 1994; Kline, 1998).

of the factor loadings were relatively large, ranging from 0.62 to 0.89, and statistically significant, and the standard errors were acceptable. Table 5 provides the standardized parameter estimates for the structural model. The squared multiple correlation (R^2) and Magnitude-of-Effect



Chi-square ($\chi^2_{(500)}$) = 1802.437 P = 0.000 CFI = 0.93 RMSEA = 0.06

Figure 3 Hypothesized SEM latent model of the relationships between state and trait cognitive and metacognitive strategy use and L2 reading test performance over time

Table 5 Standardized parameter estimates for hypothesized model regarding the state–trait reading strategy use relationships, stability of state–trait reading strategy use and its relationships to L2 reading performances across time

Observed variable	Variable	Direct effect	Error/ Disturbance	R-squared (R^2)	f^2 index
TPLN1	V1 =	0.835*F1	0.550 E1	0.698	2.311
TMON1	V2 =	0.736*F1	0.677 E2	0.541	1.178
TEVA1	V3 =	0.884*F1	0.467 E3	0.782	3.587
TCOM1	V4 =	0.846*F2	0.534 E4	0.715	2.508
TMEM1	V5 =	0.850*F2	0.526 E5	0.723	2.610
TRET1	V6 =	0.872*F2	0.489 E6	0.761	3.184
SPLN1	V7 =	0.814*F3	0.581 E7	0.663	1.967
SMON1	V8 =	0.685*F3	0.728 E8	0.470	0.886
SEVA1	V9 =	0.887*F3	0.462 E9	0.787	3.694
SCOM1	V10 =	0.836*F4	0.549 E10	0.699	2.322
SMEM1	V11 =	0.806*F4	0.592 E11	0.650	1.857
SRET1	V12 =	0.874*F4	0.485 E12	0.764	3.237
GRLEX1	V13 =	0.731*F5	0.682 E13	0.534	1.145
GRLEX2	V14 =	0.783*F5	0.621 E14	0.614	2.146
GRLEX3	V15 =	0.845*F5	0.535 E15	0.714	2.496
TxtCOM1	V16 =	0.721*F6	0.693 E16	0.520	1.083
TxtCOM2	V17 =	0.860*F6	0.510 E17	0.740	2.846
TPLN2	V18 =	0.852*F7	0.524 E18	0.725	2.636
TMON2	V19 =	0.757*F7	0.653 E19	0.573	1.341
TEVA2	V20 =	0.896*F7	0.443 E20	0.804	4.102
TCOM2	V21 =	0.880*F8	0.476 E21	0.774	3.424
TMEM2	V22 =	0.820*F8	0.572 E22	0.672	2.048
TRET2	V23 =	0.873*F8	0.488 E23	0.762	3.201
SPLN2	V24 =	0.828*F9	0.560 E24	0.686	2.184
SMON2	V25 =	0.739*F9	0.674 E25	0.546	1.202
SEVA2	V26 =	0.895*F9	0.447 E26	0.800	4.000
SCOM2	V27 =	0.834*F10	0.552 E27	0.695	2.278
SMEM2	V28 =	0.812*F10	0.583 E28	0.660	1.941
SRET2	V29 =	0.827*F10	0.563 E29	0.683	2.154
GRLEX4	V30 =	0.622*F11	0.783 E30	0.387	0.631
GRLEX5	V31 =	0.763*F11	0.647 E31	0.581	1.386
GRLEX6	V32 =	0.792*F11	0.611 E32	0.627	1.680
TxtCOM3	V33 =	0.786*F12	0.618 E33	0.618	1.617
TxtCOM4	V34 =	0.765*F12	0.644 E34	0.586	1.415
TCSU1	F2 =	0.947*F1	0.321 D2	0.897	8.708
SMSU1	F3 =	0.765*F1	0.644 D3	0.585	1.049
SCSU1	F4 =	0.218*F2 0.759*F3	0.370 D4	0.863	6.299
EFL LEX-GR1	F5 =	0.552*F4	0.834 D5	0.305	0.438
EFL RCOM1	F6 =	0.993*F5	0.117 D6	0.986	70.428
TCSU2	F8 =	0.956*F7	0.293 D8	0.914	10.627
SMSU2	F9 =	0.788*F7	0.616 D9	0.621	1.638
SCSU2	F10 =	0.749*F9 0.255*F8	0.295 D10	0.913	10.494
EFL LEX-GR2	F11 =	0.396*F10 0.575* F5	0.649 D11	0.579	1.375

(*continued*)

Table 5 (*continued*)

Observed variable	Variable	Direct effect	Error/ Disturbance	R-squared (R^2)	f^2 index
EFL-RCOM2	F12 =	0.822*F11 0.146*F6	0.384 D12	0.853	5.802

F1 = Trait metacognitive strategy use Time 1; **F2** = Trait cognitive strategy use Time 1, **F3** = State metacognitive strategy use Time 1; **F4** = State cognitive strategy use Time 1, **F5** = EFL Lexico-Grammatical ability Time 1; **F6** = EFL reading comprehension ability Time 1, **F7** = Trait metacognitive strategy use Time 2; **F8** = Trait cognitive strategy use Time 2, **F9** = State metacognitive strategy use Time 2; **F10** = State cognitive strategy use Time 2, **F11** = EFL Lexico-Grammatical ability Time 2; **F12** = EFL reading comprehension ability Time 2.

(ME) estimates[5] (f^2) for each variable are included in Table 5. The R^2 statistic is an index for determining the amount of variance accounted for by the predictor variables. For the purpose of this study, f^2 (Cohen, 1992: 175) was computed as effect size (ES) indices[6] for R^2.

3 Structures of the SEM measurement models

In Figure 3, MSU was composed of planning, monitoring and evaluating strategies, whereas CSU was composed of comprehending, memory and retrieval strategies. The structures of these variables were similar to those tested by Purpura (1999). The structures of the EFL reading test performance were the same on both occasions. The reading comprehension test performance was explained by not only the lexico-grammatical reading ability, but also the ability to comprehend texts, such as understanding main ideas and making inferences.

4 Answers relevant to RQ1 (To what extent does trait MSU at Time 1 relate to trait MSU at Time 2?)

Figure 3 shows that the correlation coefficient between trait MSU at Time 1 and trait MSU at Time 2 was also 0.56 ($R^2 = 0.31$; $f^2 = 0.45$,

[5] Magnitude-of-Effect (ME) estimates require careful interpretation since findings that are statistically significant are not always meaningful in a practical sense (i.e., they may sometimes be an artifact of the sample size). Use of ME estimates can assist the researcher in establishing whether the statistically significant findings leading to rejection of the null hypothesis (H_0) are of practical or meaningful significance within the context of an empirical investigation.

[6] The formula to compute f^2 is: $R^2/1 - R^2$. The following are interpretations of ES: $f^2 = 0.02$ as small effect size; $f^2 = 0.15$ as medium effect size; $f^2 = 0.35$ as large effect size. Medium effect size represents an effect likely to be visible to the naked eye of a careful observer, whereas small effect size is set to be noticeably smaller than medium effect size but not so small as to be trivial. Large effect size is set to be the same distance above medium as small was below it.

large ES). This finding suggests that 31% was shared between trait MSU at Times 1 and 2. A separate SEM analysis of the relationship between trait CSU at Times 1 and 2 was also 0.56 ($R^2 = 0.31$; see Appendix C). The finding suggests that even a trait construct that is believed to be quite stable is changing over time. Perhaps in the context of L2 learning, a trait construct can be unstable, particularly where there is room to improve language proficiency. Future research needs to search for logical explanations as to why the correlation between the same trait measured over time could be low (apart from measurement limitations). It may well be that in L2 learning contexts, strategic competence develops over time in the way that interlanguage does (Gass & Selinker, 2001). In this study, trait MSU at Time 2 is a more current version than trait MSU at Time 1. Furthermore, the low consistency of trait MSU found in this study might be related to the context of the study which was situated within a formal classroom setting where instruction would be expected to improve L2 learning. Given the present evidence, generalizations must be made with caution when employing a trait strategy use measure in a cross-sectional study.

5 Answers relevant to RQ2 (To what extent does trait MSU affect trait CSU?)

In Figure 3 it can be seen that ßs from trait MSU to trait CSU were 0.95 ($R^2 = 0.90$; $f^2 = 9.0$, large ES) for Time 1 and 0.96 ($R^2 = 0.92$; $f^2 = 11.50$, large ES) for Time 2. It was found that knowledge about how one plans, monitors and evaluates in general exerts an executive (higher-order) influence on knowledge about how one goes about processing information when using the language (at a lower-order level). These findings are in line with the previous research discussed above and validate Bachman and Palmer's (1996) model of strategic competence in regard to the metacognitive function of human cognition. The present findings were similar to Purpura's (1999) who found that MSU had a direct, strong effect on CSU. The present study, however, provides two observations of such effects which were highly consistent over time.

6 Answers relevant to RQ3 (To what extent does trait CSU affect state CSU?)

As indicated in Figure 3, it was found that ßs from trait CSU to state CSU were 0.22 ($R^2 = 0.04$; $f^2 = 0.04$, small ES) for Time 1 and 0.25 ($R^2 = 0.06$; $f^2 = 0.06$, small ES) for Time 2. Trait CSU cannot predict much of the degree of state CSU (only about 4–6% in this study).

There are two plausible explanations for this finding. First, it is conceivable that trait CSU does not have an 'executive function' (as trait MSU does), and thus how state CSU is to be used may largely depend on trait and state MSU (as discussed further below). It might also be that the direct influence of trait MSU on trait CSU (as discussed in answers to RQ1) was responsible for the relationship found in this model. Perhaps trait CSU is state-like in character, suggesting that it may be difficult to see cognitive strategy use as trait (i.e., as context-free). Second, it may also be the case that the extent to which one actually uses cognitive strategies depends on the context in which the language use occurs (e.g., degrees of complexity of texts, test questions and time constraints; see Robinson, 2001). Thus, some contexts may require individuals to infer meanings more than to summarize and identify main ideas.

7 Answers relevant to RQ4 (To what extent does trait MSU affect state MSU in a specific context?)

As can be seen in Figure 3, ßs from trait MSU to state MSU on both occasions suggested a strong and stable relationship between trait MSU and state MSU. The ßs were 0.76 ($R^2 = 0.58$; $f^2 = 1.38$, large ES) for Time 1 and 0.79 ($R^2 = 0.62$; $f^2 = 1.63$, large ES) for Time 2. The ßs show that the relationship between trait and state MSU was positive and stable over time. That is, unlike CSU, what individuals generally perceived themselves to be *doing* metacognitively was statistically highly related to what they actually *did* in the specific language use situations. Nevertheless, the degree to which trait MSU accounted for state MSU appeared *moderate*, suggesting that trait MSU and state MSU are different in nature. An examination of the values of the R^2 suggests that trait MSU predicted state MSU only up to about 62% (large ES), implying again that 'context' may determine the degree of state MSU.

8 Answers relevant to RQ5 (To what extent does state MSU affect state CSU?)

Inspection of Figure 3 suggests a strong relationship between state MSU and state CSU. ßs were 0.76 ($R^2 = 0.58$; $f^2 = 1.38$, large ES) for Time 1 and 0.75 ($R^2 = 0.56$; $f^2 = 1.27$, large ES) for Time 2. This finding suggests that there was 56–58% of shared variance between state MSU and state CSU. The regression coefficients show that the relationships were strong (large ES) and consistent over time. The findings regarding the executive roles of state MSU in relation to state CSU is consistent with the findings on the executive roles of trait

MSU in relation to trait CSU (as reported for RQ2). The findings imply that during actual information processing, online metacognitive strategies work closely with online cognitive strategies to achieve the learner's communicative goals. Hence, in line with Purpura (1999), neither cognitive strategies nor metacognitive strategies alone lead to performance success; rather both do. Nevertheless, the ßs in Figure 3 suggest that cognitive strategies (as related to the language processes of understanding, learning or using) are not the same process as metacognitive strategies (as related to the regulation of such processes). During information processing, cognitive strategies may be operating independently not only of each other, but also of any metacognitive strategies (Rumelhart, 1977) and hence their independent operations would decrease the strength of the relationship not only among cognitive strategies, but also between cognitive and metacognitive strategies.

9 Answers relevant to RQ6 (To what extent does state CSU affect a specific language test performance?)

Data presented in Figure 3 indicate that the ßs of state CSU on lexico-grammatical performance were 0.55 ($R^2 = 0.30$; $f^2 = 0.42$, large ES) for Time 1 and 0.40 ($R^2 = 0.16$; $f^2 = 0.19$, medium ES) for Time 2. That is, state CSU at Time 1 explained 30% of the lexico-grammatical performance variance (EFL LEX-GR), whereas at Time 2, it explained only 16% of the test score variance. It should be noted that the ß in the full SEM model and the ß in the cross-sectional SEM model at Time 2 (see Appendix B) differ significantly. In the cross-sectional analysis, the ß was 0.61 ($R^2 = 0.37$; $f^2 = 0.59$, large ES) which was similar to that at Time 1. It is plausible that as Purpura (1999) suggests, in the full SEM model, the simultaneous analysis of more information from different data sets yields changes in parameter estimates. Since the ß in the full model decreased dramatically, results from the IRT test analyses for the mid-term and final tests and the test mean scores were further examined. It was found that the mid-term test might have been less difficult than the final test. SEM might have detected differences in the two test performances when analyzing the data simultaneously resulting in a lower ß. Given the present finding, it may be argued that as Purpura (1999) rightly pointed out, the degree of strategy use does not necessarily result in better performance. If the task demand requires language ability beyond that which individuals possess, strategy use cannot always overcome the difficulty. Hence, as found in this study, only 16% of the test score variance was explained by state CSU at Time 2.

In Figure 3, the indirect ßs of state CSU and reading comprehension performance at Times 1 and 2 were 0.54 (i.e., 0.55×0.99; $R^2 = 0.29$; $f^2 = 0.41$, large ES) and 0.33 (i.e., 0.40×0.82; $R^2 = 0.11$; $f^2 = 0.12$, medium ES). Lexico-grammatical ability was found to influence reading comprehension performance (EFL RCOM) such as finding main ideas and specific information and making inferences. It was found that lexico-grammatical ability was a strong predictor of reading comprehension ability (ß = 0.99 for Time 1 and ß = 0.82 for Time 2). As discussed by Purpura (1999, p. 120), this finding is not surprising because the test-takers at the intermediate levels 'depended more heavily on their ability to decode text at the lexical and syntactic levels, rather than at inferential levels'. Since performance on the rational cloze (Section 1) depends on the skills of using contextual cues as well as syntactic cues to supply missing words, these skills could predict the ability to read text from details, main ideas and inferences (Section 2). In addition to this, the ß between EFL - LEXGR1 and EFL LEX-GR2 was 0.57 ($R^2 = 0.32$; $f^2 = 0.47$, large ES) and the ß between EFL RCOM1 and EFL RCOM2 was 0.15 ($R^2 = 0.02$; $f^2 = 0.02$, small ES). The regression coefficients suggest that lexico-grammatical performance might be more stable over time (as this knowledge remains quite permanent in LTM) than reading comprehension performance (as text and task difficulty and demands change according to the context).

Table 6 presents the decomposition of the direct and indirect effects of state-trait MSU and CSU on the EFL test performance at Times 1 and 2. Based on Table 6, state and trait cognitive and metacognitive strategy use can be seen as a large, complex system of behaviors that are highly interrelated with each other and the clusters of state and trait strategy use had a significant, differential (direct or indirect), positive effect on EFL reading test performance. In Table 6, state CSU was shown to be exerting mediating effects of trait MSU, trait CSU and state MSU on language test performance. The total effects suggest differences in the awareness of strategy use by individuals with different achievement levels. That is, more successful test-takers reported higher awareness of state and trait cognitive and metacognitive strategy use than less successful ones. The present findings support not only previous research on the L2 reading strategies which affect reading performance (e.g., Barnett, 1988; Block, 1992; Carrell, 1998), but also Bachman and Palmer's (1996) communicative

Table 6 Decomposition of the total effects of EFL test performances in Times 1 and 2

Variable	Direct and indirect effect
EFL LEX-GR1 (F5)	0.435 F1(TMSU1) + 0.121 F2 (TCSU1) + 0.419 F3 (SMSU1) + **0.552** *F4 (SCUS1) + 0.039 D2 + 0.270 D3 + 0.204 D4 + 0.834 D5
EFL RCOM1 (F6)	0.432 F1 (TMSI1) + 0.120 F2 (TCSU1) + 0.416 F3 (SMSU1) + 0.548 F4 (SCSU1) + **0.993** *F5(EFL LEX-GR1) + 0.038 D2 + 0.268 D3 + 0.203 D4 + 0.828 D5 + 0.117 D6
EFL LEX-GR2 (F11)	0.250 F1 (TMSU1) + 0.069 F2 (TCSU1) + 0.241 F3 (SMSU1) + 0.317 F4 (SCSU1) + **0.575** *F5 (EFL LEX-GR1) + 0.330 F7 (TMSU2) + 0.101 F8 (TCSU2) + 0.296 F9 (SMSU2) + **0.396** *F10 (SCSU2) + 0.022 D2 + 0.155 D3 + 0.117 D4 + 0.479 D5 + 0.030 D8 + 0.183 D9 + 0.117 D10 + 0.649 D11
EFL RCOM2 (F12)	0.268 F1(TMSU1) + 0.074 F2 (TCSU1) + 0.259 F3 (SMSU1) + 0.341 F4 (SCSU1) + 0.617 F5(EFL LEX-GR1) + **0.146** *F6 (EFL RCOM1) + 0.271 F7 (TMSU2) + 0.083 F8 (TCSU2) + 0.244 F9 (SMSU2) + 0.326 F10 (SCSU2) + **0.822** *F11 (EFL LEX-GR2) + 0.024 D2 + 0.167 D3 + 0.126 D4 + 0.514 D5 + 0.017 D6 + 0.024 D8 + 0.150 D9 + 0.096 D10 + 0.533 D11

Parameter estimates in bold indicate direct effects.

ability model which takes non-linguistic factors into account in explaining language test performance. This study also lends support to the interactionalist perspective which views language test performance as a result of an interaction among traits, contextual features and strategic competence (Chapelle, 1998).

V Metadiscussion

As evident in this study, strategic competence as a non-linguistic factor in L2 use is highly complex, hierarchical and multi-dimensional. Strategic competence can be understood via an examination of individuals' knowledge of how they generally perceive using a set of strategies over a variety of contexts and their reported actual use of the strategies in a specific context. The cognitive strategy construct consisted of a set of comprehending, memory and retrieval strategies, whereas the metacognitive strategy construct is composed of a set of planning, monitoring and evaluating strategies. In line with Purpura (1999), the present study suggests that state and trait metacognitive strategies exert an executive function over cognitive strategy use (see RQs 1–5). State CSU is in turn directly related to the specific test performance (see RQ 6). The present study contributes to the theory of strategic competence formulated by Bachman and Palmer (1996) in that two strategic competence facets need to be recognized: *strategic knowledge* (knowledge about what, how and when to use a set of strategies) and *strategic regulation* (online realization and regulation of cognitive processing, particularly when problems or difficulties arise).

1 Strategic competence and language performance

The question of how much strategy use can explain language test performance variance arises. Previous research and this study repeatedly found that MSU and CSU are *weakly to moderately related* to language performance (up to 30%), though in some contexts, 'strategic processing may contribute to a language test performance more than others' (Bachman, 1990, p. 105). There are a number of plausible reasons for the limited strength of this relationship. The first reason we put forward is that in language use, such as test-taking, factors such as communicative language ability, test-method facets, test-taker characteristics and random factors will simultaneously affect language use performance (Bachman & Palmer, 1996). Since strategic competence is only part of communicative language ability, one should not expect it to account for language performance to a large extent. Furthermore, according to Anderson (2005), success in using strategies may depend on certain conditions: (1) whether the strategy relates well to the L2 task at hand; (2) whether the strategy is linked well with other strategies and processes relevant to the given task; and (3) whether the strategy coordinates well with the learner's learning style. Perhaps strategic competence by L2 learners is in the process of development and hence incomplete (see RQ1). Phakiti (2005) examined test-takers' calibration (accurate ability to judge one's performance success) and found that test-takers in general were overconfident in their test performance. Most relevant to the present study is that they were found to be overconfident about difficult questions but underconfident about easy questions. These findings suggested that test-takers might have lacked accurate self-assessment and evaluation, which in turn may have resulted in poor strategy use.

The second possible explanation for the weak to moderate correlations between strategy use and test performance is that success in a language learning or use depends on a number of cognitive and noncognitive factors, including language competence (Bachman & Palmer, 1996), the levels of linguistic thresholds in a particular context of language use (Bernhardt, 2000), individuals' working memory capacity (Robinson, 2001), levels of accurate automaticity in language processing (Segalowitz, 2003), motivation to use the language

(Dörnyei, 2001), and task difficulty, demands and constraints (Skehan, 1998). Given that these factors account for language performance as well, the amount of variance that remains to be shared between strategic competence and test performance is necessarily limited.

The third explanation that can be offered is that much of the overall information processing during language use involves lower-level processing (e.g., automatic, unreflective and unconscious processing, especially in easy, familiar tasks or environment; see Sternberg, 1998) and higher-level processing (e.g., conscious awareness of what and how they are using the language). Some apparently metacognitive processes that are embedded within the automatic processes are not strictly speaking metacognitive although they have a control function (Gagné *et al*., 1993). Higher-level processing is apparent when individuals experience difficulty or aim to achieve a specific, hard-to-achieve goal. Therefore, there is a constant back-and-forth relationship between automatic and strategic processing, which contributes a varying degree of a relationship to a specific language performance.

The fourth possible reason is that there remain methodological and analytical limitations in researching strategic competence. Methodologically, researchers are faced with how to connect their measures with the constructs of interest (see Tseng *et al*., 2006). Strategic competence, like many other constructs, does not often lend itself to direct observation and hence to understand it, researchers are dependent upon learners' potentially fallible accounts of their mental knowledge and processes. Though structural equation modeling is a powerful analytical tool, variance-based modeling is not necessarily a window to any particular mind because individuals uniquely vary among each other and human behaviors are dynamic, situational and contextual (Phakiti, 2008). Since human thinking is not open to direct scrutiny, structural models from empirical data only allow us to speculate about what may be happening inside the mind.

2 Limitations and implications

There are limitations to the present study that are worth mentioning because they have implications for further research. First, the findings may be in part an artifact of the research instruments (i.e., multiple-choice reading tests and Likert-scale questionnaires), and the setting in which the study was carried out. Moreover, this study is limited to EFL reading test performance and hence generalizations of the findings to other language skills are limited (see VanPatten, 1994).

Multiple-choice tests are also limited in their ability to assess authentic language use and hence performance based on this test technique may not be the best indicator of an individual's language proficiency. The strategy questionnaires used for this study may likewise have failed to document the full array of strategies test-takers used both habitually and on the test in question. In addition, the sample for the study was limited to Thai EFL university students. Hence, an implication for further research is to examine the relationship between strategic competence and language test performance in other language skills, using a variety of test formats and more comprehensive strategy questionnaires in various EFL and ESL (English as a second language) and languages other than English settings, thereby allowing future cross-cultural comparisons.

Second, this study was confined to observations from two test occasions. An ideal scenario would be to gather a series of data sets documenting state and trait cognitive and metacognitive strategy use and high-stakes test performance at various time points from the same learners. It should be noted, however, that a longitudinal quantitative study such as the one reported here is quite difficult to execute. Retaining the same group of learners over a long period of time can be expensive, time-consuming and impractical in many real-world testing contexts.

Third, the SEM model was not applied to determine whether the factorial structure between each group of, for example, successful and unsuccessful test-takers, and males and females is invariant and hence claims about group-specific models cannot be made. Therefore, future studies could build on the current findings by employing multi-group analyses (see Purpura, 1999).

VI Concluding remarks

Previous research and the present study show that all learners/test-takers engage in a range of strategic behaviours; however, they also show that such strategic activities, often prolonged and varied, do not necessarily lead to language success. It may be that optimal strategic behaviours differ under different conditions. Future research needs to investigate (1) typical situations that inhibit strategy use and (2) those that are most conducive to its optimal use in language learning and/or language performance. Research on strategic competence is a dynamic area of study where new paths and perspectives are still unfolding. The present study has offered some further insights into our conceptualization of strategic competence and its role in L2 communicative language ability but there is much work still to be done. As McNamara (1996, p. 88) notes, 'progress in understanding the nature of second language performance testing and conducting the necessary research requires a commitment to rigorous analysis of what is involved in a language test performance'.

Acknowledgements

I would like to thank the following people who have played metacognitive roles during the completion of this research project and earlier drafts of this paper: Brian Lynch, Tim McNamara, Paul Gruba, Jim Purpura, Antony Kunnan, Fred Davidson, Andrew Cohen, Carsten Roever, John Read, Brian Paltridge, Sirikarn Chanruang, Marie Stevenson, Lindy Woodrow, David Hirsh, Diana Montgomery, the EQS technical support team, the three anonymous reviewers and the Editors.

VII References

Adams, R. J. & **Khoo, S-K.** (1996). *Quest* (Version 2.1) [Computer software]. Victoria, Australia: Australian Council for Educational Research.

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Alexander, P. A., Schallert, D. L. & **Hare, V. C.** (1991). Coming to term: How researchers in learning and literacy skills talk about knowledge. *Review of Educational Psychology*, 61, 315–343.

Anderson, J. R. (2000). *Learning and memory: An integrated approach* (2nd ed.). New York: John Wiley.

Anderson, N. J. (1999). *Exploring second language reading: Issues and strategies*. Boston, MA: Heinle and Heinle.

Anderson, N. J. (2005). L2 learning strategies. In E. Hinkel, editor, *Handbook of research in second language teaching and learning* (pp. 757–771). Mahwah, NJ: Lawrence Erlbaum.

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

Bachman, L. F. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing*, 17, 1–42.

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

Baker, L. & **Brown, A. L.** (1984). Metacognitive skills and reading. In P. D. Pearson, editor, *Handbook of reading research* (pp. 353–394). New York: Longman.

Barnett, M. A. (1988). Reading through context: How real and perceived strategy use affects L2 comprehension. *Modern Language Journal*, 72, 150–162.

Bentler, P. M. (1985–2006). *EQS Version 6 for Windows* [Computer software]. Encino, CA: Multivariate Software.

Bentler, P. M. (2006). *EQS structural equation program manual*. Encino, CA: Multivariate Software.

Bernhardt, E. B. (2000). Second language reading as a case study of reading scholarship in the 20th century. In M. Kamil & P. Mosenthal, P. D. Pearson, & R. Barr, editors, *Handbook of reading research*, Vol. III (pp. 791–811). Hillsdale, NJ: Lawrence Erlbaum.

Block, E. L. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly*, 26, 319–343.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks, CA: Sage Publications.

Campbell, D. T. & **Fiske, D. W.** (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.

Carrell, P. L. (1998). Can reading strategies be successfully taught? *Annual Review of Applied Linguistics*, 21, 1–20.

Carrell, P. L., Gajdusek, L. & **Wise, T.** (1998). Metacognition and EFL/ESL reading. *Instructional Science*, 26, 97–112.

Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen, editors, *Interfaces between second language acquisition and language testing research* (pp. 32–70). Cambridge: Cambridge University Press.

Cohen, A. D. (1998). *Strategies in learning and using a second language*. London and New York: Longman.

Cohen, A. D. (2007). The coming of age of research on test-taking strategies. In J. Fox, M. Weshe, D. Bayliss, L. Cheng, C. Turner, & C. Doe, editors, *Language testing reconsidered*. Ottawa: Ottawa University Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.

Dörnyei, Z. (2001). *Teaching and researching motivation*. Harlow: Longman.

Ellis, R. (1994). *The study in second language acquisition*. Oxford: Oxford University Press.

Flavell, J. H. (1985). *Cognitive development*. Englewood Cliffs, NJ: Prentice-Hall.

Gagné, E. D., Yekovich, C. W. & **Yekovich, F. R.** (1993). *The cognitive psychology of schooling learning*. New York: HarperCollins.

Gass, S. & **Selinker, L.** (2001). *Second language acquisition: An introductory course* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Hsiao, T-Y. & **Oxford, R. L.** (2002). Comparing theories of language learning strategies: A confirmatory factor analysis. *Modern Language Journal*, 86, 368–383.

Hudson, T. (2007). *Teaching second language reading*. Oxford: Oxford University Press.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes, editors, *Sociolinguistics: Selected readings* (pp. 269–293). Harmondsworth, Middlesex: Penguin.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.

MacCallum, R. C. & **Austin, J. T.** (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51, 201–226.

McNamara, T. (1996). *Measuring second language performance*. London and New York: Longman.

Maruyama, G. M. (1998). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage Publications.

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

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

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

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

Phakiti, A. (2003a). A closer look at gender differences in strategy use in L2 reading. *Language Learning*, 53, 649–702.

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

Phakiti, A. (2005). An empirical investigation into the nature of and factors affecting test takers' calibration within the context of an English Placement Test (EPT). *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, 3, 27–46.

Phakiti, A. (2008). Strategic competence as a fourth-order factor: A structural equation modeling approach. *Language Assessment Quarterly,* 5, 20–42.

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

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Robinson, M. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22, 27–57.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic & P. Rabbitt, editors, *Attention and performance*, Vol. 6, (pp. 573–603.). New York: Academic Press.

Schmidt, F. L. & **Hunter, J. E.** (1999). Theory testing and measurement error. *Intelligence*, 27, 183–198.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt, editor, *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu: University of Hawai'i Press.

Schmidt, R. (2001). Attention. In P. Robinson, editor, *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Schraw, G. & **Moshman, D.** (1995). Metacognitive theories. *Educational Psychology Review*, 7, 351–371.

Segalowitz, N. (2003). Automaticity and second language development. In C. Doughty & M. Long, editors, *Handbook of second language acquisition* (pp. 382–408). Oxford: Blackwell.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Song, X. (2004). Language learning strategy use and language performance for Chinese learners of English. Unpublished master's thesis. Ontario: Queen's University, Kingston, Canada.

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, 1–26.

Spielberger, C. D. (1972). Conceptual and methodological issues in anxiety research. In C. D. Speilberger, editor, *Anxiety: Current trends in theory and research*, Vol. 2 (pp. 481–493). New York: Academic Press.

Spolsky, B. (1995). *Measured words*. Oxford: Oxford University Press.

Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science*, 26,

Tseng W., Dörnyei, Z. & **Schmitt, N.** (2006). A new approach to assessing 127–140. strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, 27, 78–102.

VanPatten, B. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features and research methodology. *AILA Review*, 11, 27–36.

Appendix A: Trait and state strategy use questionnaires

Trait reading strategy questionnaire

Directions: Read each statement and indicate how you normally think when read in English. Choose 0 (never), 1 (rarely), 2 (sometimes), 3 (often), 4 (usually) or 5 (always) on each statement that best describes how you think.

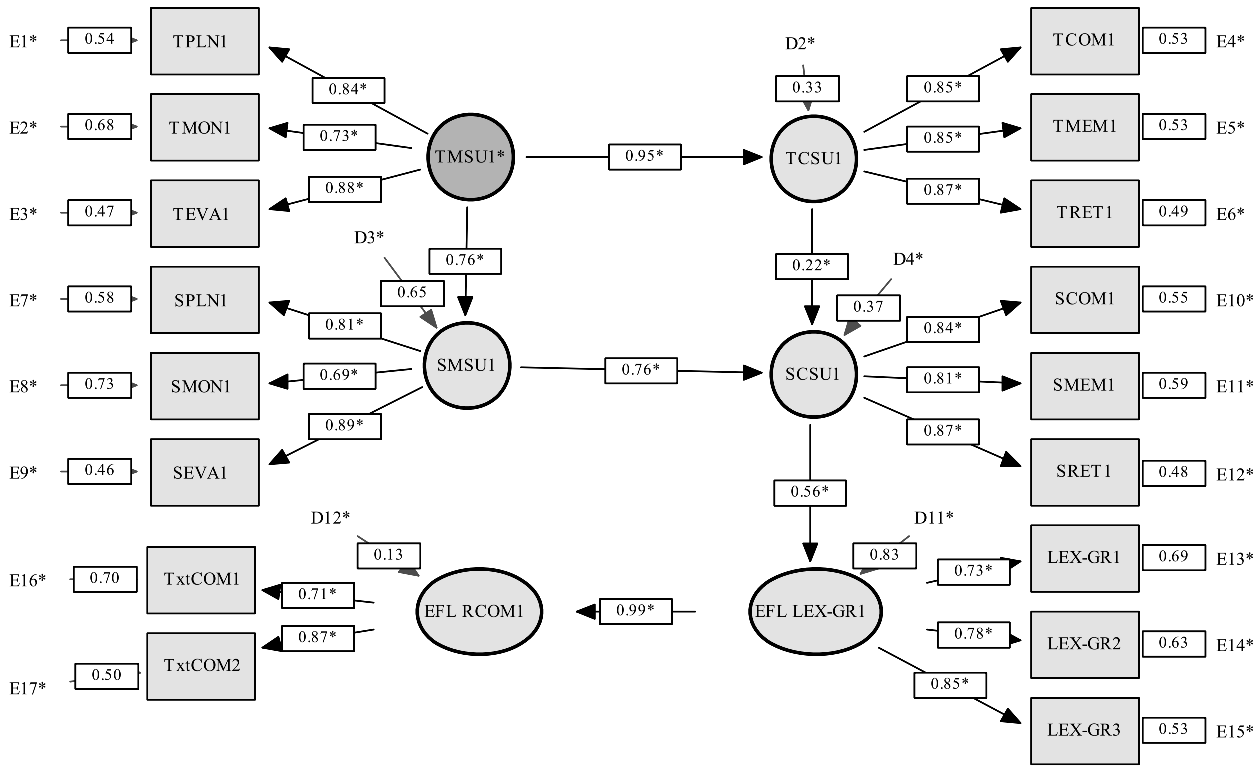
No.	Your thinking	0 1 2 3 4 5
1.	I plan what to do before I begin to read English texts or tasks.	
2.	I make sure I clarify the goals of the reading tasks.	
3	I consider steps needed to complete the reading.	
4	I make sure I understand what has to be done and how to do it.	
5.	I know what to do if my plans do not work efficiently.	
6.	I scan through the reading and tasks before I actually begin doing it.	
7.	I try to understand the relationships between ideas in the text.	
8.	I try to understand the content without looking up every word.	
9.	I think what is going to happen next while reading.	
10.	I analyze what the author means or tries to say.	
11.	I try to interpret hidden ideas/meanings in texts.	
12.	I translate text or reading tasks into my first language.	
13.	I summarize the main information in the text.	
14.	I relate the information from the reading or tasks to my prior knowledge or experience.	
15.	I reread texts or tasks several times when I feel I do not understand them.	
16.	I know which information is more or less important.	
17.	I identify or guess meanings of unknown words using context clues.	
18.	I apply my learned grammar rules while reading or completing reading tasks.	
19.	I guess meanings of unknown words using root words.	
20.	I am aware of time limitations and constraints.	
21.	I know how much the reading and tasks remain to be done while reading.	
22.	I check if I understand the text or task.	
23.	I check my own performance and progress as I move along the reading tasks.	
24.	I know when I lose concentration while reading.	
25.	I evaluate my plans or goals of my reading constantly.	
26.	I know when I should read more quickly or carefully.	
27.	I double-check my reading comprehension or performance.	
28.	I immediately correct my misunderstanding or mistakes in reading tasks when found.	
29.	I notice when and where I am confused in the text.	
30.	I know when I feel worried, tense or unmotivated while reading.	

State reading strategy questionnaire

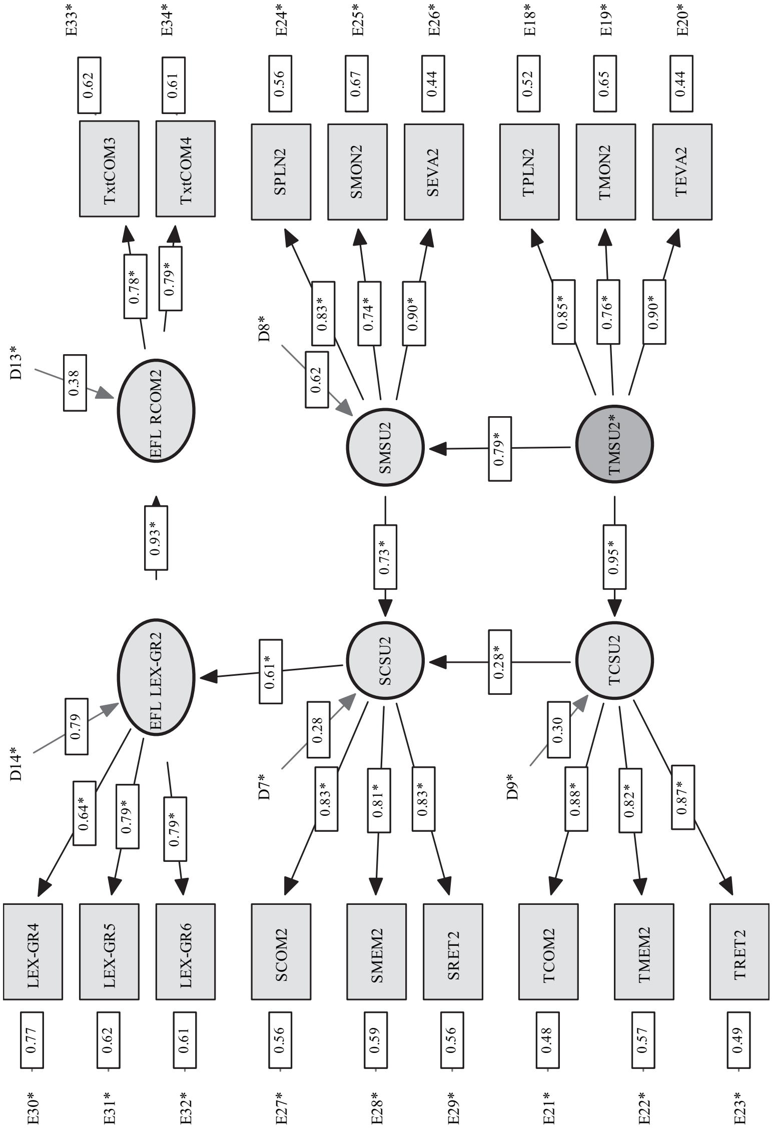
Directions: Read each statement and indicate how you actually thought when taking this reading test. Choose 0 (never), 1 (rarely), 2 (sometimes), 3 (often), 4 (usually) or 5 (always) on each statement that best describes how you thought.

No.	Your thinking	0 1 2 3 4 5
1.	I planned what to do before I began to complete this reading test.	
2.	I made sure I clarified the goals of the reading test tasks.	
3.	I considered essential steps needed to complete the reading test.	
4	I made sure I understood what had to be done and how to do it.	
5.	I knew what to do if my intended plans did not work efficiently while completing this reading test.	
6.	I flipped through the reading test before I actually stated to complete it.	
7.	I tried to understand the relationships between ideas in the text and tasks.	
8.	I tried to understand the content of the text and tasks without looking up every word.	
9.	I thought what was going to happen next while I was reading the text.	
10.	I analyzed what the author meant or tried to say in the text.	
11.	I tried to interpret hidden ideas/meanings in the texts.	
12.	I translated the text, tasks or questions into my first language.	
13.	I summarized the main information in the text.	
14.	I related the information from the text or tasks to my prior knowledge or experience.	
15.	I reread texts or tasks several times when I felt I did not understand them.	
16.	I knew which information was more or less important.	
17.	I identified or guessed meanings of unknown words using context clues.	
18.	I applied my learned grammar rules while reading and completing the reading tasks.	
19.	I guessed meanings of unknown words using root words.	
20.	I was aware of the time limitations and constraints in this test.	
21.	I knew how much of the reading and test tasks remained to be done while taking the test.	
22.	I checked if I understood the texts and reading tasks.	
23.	I checked my own performance and progress as I moved along the test tasks.	
24.	I knew when I lost concentration while completing this test.	
25.	I evaluated my plans or goals of my reading tasks constantly.	
26.	I knew when I should read or complete the test more quickly or carefully.	
27.	I double-checked my reading comprehension or performance.	
28.	I immediately corrected my misunderstanding or performance mistakes when found.	
30.	I knew when I felt worried, tense or unmotivated to complete this reading test.	

Appendix B: Cross-sectional hypothesized SEM models (Times 1 and 2)

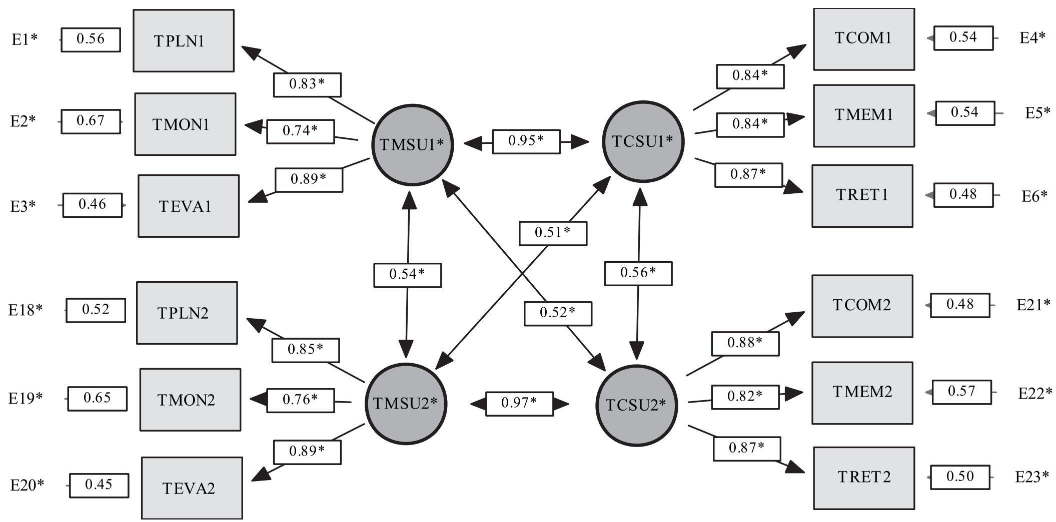


Chi-square ($\chi^2_{(107)}$) = 648.387 P = 0.00 CFI = 0.93 RMSEA = 0.09



Chi-square ($\chi^2_{(107)}$) = 587.586 P = 0.00 CFI = 0.93 RMSEA = 0.09

Appeneix C Correlations between trait metacognitive and cognitive strategy use in Times 1 and 2



Chi-square ($\chi^2_{(44)}$) = 208.037 P = 0.00 CFI = 0.97 RMSEA = 0.08