

# Mapping the fluctuating effect of strategy use ability on English reading performance for nursing students: A multi-layered moderation analysis approach

Language Testing  
2020, Vol. 37(2) 280–304  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0265532219893384  
journals.sagepub.com/home/ljt



**Yuyang Cai** 

Shanghai University of International Business and Economics, China

**Antony John Kunnan**

University of Macau, China

## Abstract

An essential hypothesis of modern language assessment theory pertains to the interaction between strategy use ability (strategic competence) and second language knowledge. However, how they interact with each other is rarely explored. Drawing on relevant research in the literature, in this paper we proposed three interaction patterns (i.e., linear, quadratic, and cuboid) in which language knowledge moderates the effect of strategy use ability on L2 reading performance. A pool of 1491 nursing students were invited to respond to three instruments, each measuring language knowledge, strategy use ability, and nursing English (L2) reading ability, respectively. Student responses were first scored using multidimensional item response theory (MIRT). Next, we applied multi-layered moderation analysis (MLMA) to these MIRT-based scores to detect the hypothetical interaction patterns. The results supported the cuboid interaction pattern or, metaphorically, the pattern of an island ridge curve (IRC). Substantially, this indicated that the effect of strategy use ability on nursing English reading performance fluctuated in a down-up-down pattern with the increase of students' language knowledge. Our study also revealed different patterns of strategy use depending on students' language knowledge level.

## Keywords

Interaction, island ridge curve, multi-layered moderation analysis, reading, strategy, threshold

---

### Corresponding author:

Yuyang Cai, School of Languages, Shanghai University of International Business and Economics, Boya Building, 1900 Wenxiang Road, Songjiang, Shanghai, China.

Email: sailor\_cai@hotmail.com

Theoretical models of language ability such as Communicative Language Ability (CLA; Bachman, 1990; Bachman & Palmer, 1996; Bachman & Palmer, 2010) and Language for Specific Purposes Ability (LSPA; Douglas, 2000) conceptualized strategic competence (i.e., metacognitive strategies) and language knowledge (i.e., lexico-grammatical knowledge, textual knowledge) as two core constituents of language ability. Both theories emphasized that successful language performance should involve the interaction between strategic competence and language knowledge. But how this interaction takes place has received insufficient theoretical and empirical attention.

In second language acquisition research, the concept of strategic competence (e.g., Phakiti, 2008a, 2008b; Purpura, 1999, 2014) is closely related to L2 learner strategy (Oxford, 1990, 2017; Oxford & Amerstorfer, 2018). Studies of L2 learner strategy started with comparing strategies used by “good” and “poor” L2 learners (e.g., Green & Oxford, 1995; Rubin, 1975, 2005). Numerous studies have shown that proficient L2 learners apply a set of learner strategies (e.g., Griffiths, 2013; Razi & Grenfell, 2012). A tempting conclusion is that successful L2 performance is linearly and positively associated with the frequency of strategy use (Jiménez, García, & Pearson, 1996). Other studies, however, have found that poor L2 learners use as many strategies as good L2 learners (e.g., Alsamadani, 2009; Gürses & Bouvet, 2016). This suggests what matters is the efficiency (doing things right) of strategy use rather than the frequency of strategy use (Grabe & Stoller, 2011; Griffiths & Inceçay, 2016; Oxford, 2017). This efficiency is believed to vary with individual learner characteristics such as L2 proficiency, motivation, and so forth (Dörnyei & Ryan, 2015). Among these individual characteristics, L2 proficiency is considered the most important (J. C. Alderson, 2000; Skehan, 1989).

Skehan (1989) claimed that strategy use might have to be “permitted” by L2 proficiency (p. 127), an idea very similar to Clarke’s (1980) “short-circuit” theory. Clarke hypothesized a critical L2 proficiency level (language threshold) that L2 learners must pass in order to have their first language skills successfully transferred to L2 use. In reality, the short-circuit theory might be too general to capture the complex interaction between strategy use and L2 proficiency. This failure is reflected in the mixed findings that did not support any such interaction (Magogwe, 2013), that supported a simple linear interaction (Sheorey & Mokhtari, 2001), or that revealed a curvilinear function (Hong-Nam & Page, 2014).

We believe that at least three reasons are responsible for these mixed findings: the unchecked measurement validity of instruments used for data collection; the exclusive focus on the “frequency” aspect of strategy use; and the limited power of the analytical techniques employed for detecting interaction (e.g., ANOVA or group-based regression). Bearing these limitations in mind, we hypothesized three patterns of moderation built on the literature and tested them with multi-layered moderation analysis (MLMA), a technique we developed for detecting various forms of interaction.

## **Brief review of the literature**

### *Learner strategy, strategic competence, and strategy use ability*

In the second language acquisition literature, learner strategy is frequently labeled as follows: *language learning strategies* to emphasize the receptive aspect of learning

(O'Malley & Chamot, 1990; Oxford, 1990); *language use strategies* to emphasize the productive aspect of learning (Cohen, 2014); and *self-regulated language learning strategies* to emphasize the intentionality of strategy use (Oxford, 2017; Oxford & Amerstorfer, 2018). Learner strategy is also referred to as *listening strategies* (Vandergrift, 2008), *speaking strategies* (Cohen, 2008), *reading strategies* (Sheorey & Mokhtari, 2001), and *writing strategies* (de Larios, Manchón, Murphy, & Marín, 2008) that are illustrative of the learner strategies in particular language skills. Moreover, they share a common feature in referring to the mental processes or behaviors that language learners employ in L2 situations (Cohen, 2014; Dörnyei & Ryan, 2015; Oxford, 2017; Purpura, 2014).

In the last decade, a move from the “frequency” towards the “competence” view of strategy has been proposed. Oxford (2017) argued that low-frequency strategy use does not necessarily mean ineffective or inefficient learning; and vice versa, high frequency does not always indicate successful learning. Tseng, Dörnyei, and Schmitt's (2006) proposal of the “capacity” (what students can do) of vocabulary learning strategies is a reflection of such a concern. Language assessment theorists such as Bachman (1990) and Bachman and Palmer (1996, 2010) conceptualized strategy use as “strategic competence” and refer to the construct as the ability to use different metacognitive strategies (e.g., goal-setting, monitoring, evaluating) during language performance. Phakiti (2008a, 2008b) re-conceptualized the construct of strategic competence as a hierarchical concept containing strategic factors at two levels: knowledge (“trait”) of metacognitive and cognitive strategies at the higher level and online use (“state”) of metacognitive and cognitive strategies at the lower level. By distinguishing “trait” from “state,” Phakiti recaptured the “competence” flavor of strategic competence.

Inspired by the literature, we use the term *strategy use ability* to re-emphasize the “competence” aspect of strategic competence. Strategy use ability, therefore, refers to L2 students' ability to deploy appropriate metacognitive strategies (planning, monitoring, and evaluating) and cognitive strategies (i.e., memory, retrieving, and comprehending) to achieve successful language performance. Strategy use ability should not be regarded as bearing a radically different connotation from the term “strategic competence” (Bachman & Palmer, 2010; Douglas, 2000; Purpura, 2014). Rather, this term should be taken as an extension of strategic competence to include cognitive strategy use (e.g., Phakiti, 2008a, 2008b).

### *The moderation of language proficiency on strategy use during L2 reading*

In learner strategy use research, the past two decades have witnessed increased attention on the variation of strategy use effect on L2 reading, depending on L2 proficiency (often represented with grammatical knowledge). Sheorey and Mokhtari's (2001) study addressed the strategy use of 152 non-native English speakers (with a mean age of 21.75 years) studying in an American university. Strategy use frequency was measured using a five-point Survey of Reading Strategies (SORS; Mokhtari & Sheorey, 2002). The SORS consisted of 28 items in three categories: metacognitive strategies (e.g., planning, goal-setting, monitoring), cognitive strategies (e.g., direct problem-solving strategies such as comprehending, guessing, rereading, problem-solving) and support strategies (e.g., memorizing, note-taking). L2 reading was measured using a six-point self-report scale

ranging from 1 (for poor) to 6 (for excellent). Students who reported three points (50% of the total score) or below were put into the low-ability group, while those who reported five or six points (at or above 63.3%) were put into the high-ability group. This resulted in 78 students in the former group and 34 in the latter. Results of a *t*-test revealed more use of metacognitive strategies and problem-solving strategies (i.e., cognitive strategies) in the high-ability group than in the low-ability group. The researchers attributed this difference to the interaction between L2 proficiency, L2 reading ability, and strategy use. As they put it, high L2 proficiency led to enhanced L2 reading ability, which again led to the increased use of reading strategies, and which again led to improved L2 reading ability. This interpretation, however, is premature for at least two reasons. First, the interpretation linked to L2 proficiency was made in the absence of L2 proficiency data in their study. Second, their study only had two data points (i.e., a low-reading-ability group and a high-reading-ability group), a limitation that did not allow the study to detect a nonlinear interaction, if such a pattern were to exist.

Hong-Nam and Page (2014) examined the variation of reading strategy use with L2 proficiency and L2 reading among 432 Korean undergraduate students studying English. Students' reading strategy use was measured using the SORS (i.e., metacognitive strategies, cognitive strategies and support strategies), and their L2 proficiency and L2 reading ability were both self-rated as beginning, intermediate or advanced. Results of an ANOVA showed that strategy use was related to L2 reading linearly. Meanwhile, the intermediate L2 proficiency group used all strategies ( $M = 3.50$ ), more than either the advanced group ( $M = 3.44$ ) or the beginning group ( $M = 3.29$ ). Although the mean differences seemed to be small, they were all statistically significant ( $F = 23.82, p = .000$ ).

This curvilinear pattern was consistent with Hong-Nam and Leavell's (2006) early findings about strategy use for L2 learners (Oxford, 1990), in which Hong-Nam and Leavell interpreted this curvilinearity as "unexpected" and a result of good learners' higher enthusiasm in strategy use. However, Hong-Nam and Page went no further to link these findings to L2 proficiency and its effect on L2 reading.

This brief review of the literature reinforced our earlier discussion about the complex interaction between strategy use and L2 proficiency during L2 reading performance. Dependent on L2 proficiency, the actual interaction can be non-significant, linear, or curvilinear. Instead of seeing these as contradictory patterns, we consider the various patterns as misleading manifestations of the real interaction owing to a lack of well-controlled conditions. These limitations include poor attention to the ability of strategy use, unfocused or even unstructured measures of language proficiency, unchecked measurement validity, and, more importantly, limited power of statistical techniques used to detect the interactions.

The current study addresses these limitations. First, instead of measuring the frequency of strategy use during their language learning and use, we asked students to self-rate the efficiency of their strategy use (i.e., their ability to deploy appropriate strategies) when accomplishing a real reading test. In this way, our study was able to capture the "ability" instead of the "frequency" aspect of strategy use. Second, in order to ensure accurate measurement of language proficiency, we referred to language knowledge, the major predictor of language proficiency according to both theoretical (Bachman & Palmer, 1996, 2010) and empirical studies (Jeon & Yamashita, 2014), and measured the construct with a lexico-grammatical test constructed as form-meaning, according to the

most current model of assessing grammar (Purpura, 2004, 2017). Third, instead of directly using the raw score totals of each measure to explore relationships, we used Multidimensional Item Response Theory (MIRT) to assess the quality of each measure and derive latent scores. MIRT has been acknowledged as one of the most appropriate psychometric models for test scoring for at least three reasons: accounting for confounding factors such as item features (difficulty, discrimination, guessing); controlling for misfit between item difficulty and person ability; and providing granular information for factors determining the correct response to the test item.

Most prominently, in our study we used multi-layered moderation analysis (MLMA), an analytical model we specially developed for addressing the complex phenomenon of interactions. As reviewed earlier, previous studies relied heavily on group comparisons of strategy use for the detection of interactions. Although this approach does provide us a view of the effect of strategy use across different L2 proficiency groups (usually two or three groups), this understanding is deemed to be limited, as it is unable to reveal the continuous change of the effect of strategy use along the continuum of L2 proficiency. Encapsulated in this approach is the issue of subjective grouping. As different studies used different grouping criteria, results yielded by different studies are hardly comparable. It is unclear whether these different patterns are distorted presentations of the same true phenomenon, or true reflections of the same phenomenon but for students of different L2 proficiency.

To search for more accurate understandings, appropriate analytical techniques must be developed to meet such a subtle need. The next part introduces the analytical approach called multi-layered moderation analysis (MLMA) that we specially developed for detecting complex moderation pattern(s).

## The multi-layered moderation analysis (MLMA)

In statistics, the idea that one predictor interferes with the predictive relation of another predictor to an outcome variable corresponds to the concept of interaction or moderation (Klein & Moosbrugger, 2000). In fact, what this moderation describes is only a linear pattern of concurrent change of two variables and, therefore, it is labeled as *linear* moderation in the current study. In order to detect more complex moderation patterns as implied in the strategy use literature, we created two more complex patterns of moderation: a *quadratic* pattern (with one curve on the concurrent change projection) and a *cuboid* pattern (with two curves on the concurrent change projection). The quadratic pattern allows the detection of a nonlinear interaction that was discovered by Hong-Nam and Leavell (2006), and the cuboid pattern is the authors' attempt to explore a more subtle pattern. As the construction of the comprehensive model is successive by nature, this approach is called multi-layered moderation analysis (MLMA).

The MLMA for the current study involved four steps: (1) testing a baseline structural equation model (SEM) with strategy use ability and language knowledge as predictors and L2 (nursing English) reading ability as the outcome variable; (2) testing the linear moderation model by adding to the baseline model a latent production term of strategy use ability and language knowledge (see Klein & Moosbrugger, 2000); (3) testing the quadratic moderation model by adding to the linear moderation model a latent production term

of strategy use ability and language knowledge squared; and (4) testing the cuboid moderation model by adding to the quadratic model a latent product term of strategy use ability and language knowledge cubed. In this way, the concurrent changes of strategy use ability effect and language knowledge can be captured by the terms involving the various moderation forms included in the comprehensive model (B. Q. Muthén, 2012).

A complete MLMA model is tested in such a way that the moderation terms are tested one at a time, moving from the lowest to the highest order. Once a higher-order moderation term is confirmed (e.g., a cuboid moderation term), the lower-order moderation terms (e.g., the quadratic moderation and the linear moderation) are automatically subsumed. In other words, the confirmation of a higher-order moderation does not necessarily reject its lower order moderation, but entails a further evolution of its previous form. Once the final form is determined, the relationship between language knowledge and the effect of strategy use ability on L2 reading ability can be specified by transforming the moderation elements in the final MLMA model (see Muthén, 2012). A more technical introduction to MLMA is available in Cai and Kunnan (2019). A linear moderation pattern, a quadratic moderation pattern, and a cuboid moderation pattern are visualized in Figure 1(a), (b), and (c), respectively.

The invention and application of the MLMA is theory driven. At the time of the current study, the authors could not find such an analytical approach that could be directly applied to tap into the nonlinear moderation pattern(s) between language knowledge and strategy use that is strongly implied in strategy research literature. In this sense, the current study was a unique study that synergized theory and methodology.

## The study

### Research questions

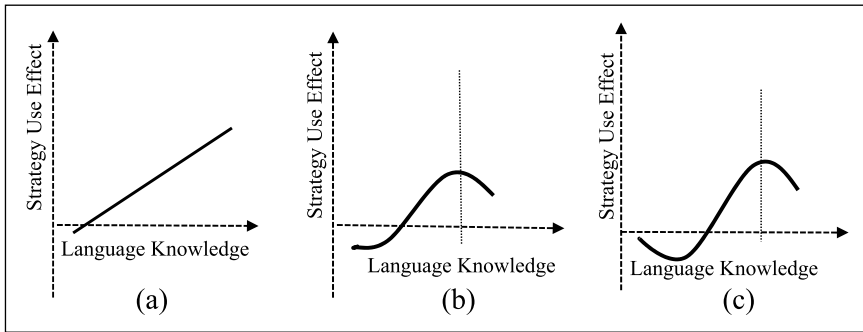
This study addressed two questions:

1. How does language knowledge moderate the relationship between strategy use ability and L2 reading ability? Is it linear, quadratic, or cuboid?
2. Does strategy use ability differ among students with different levels of language knowledge? If it does, to what extent do they differ?

The answer to the first question would help to reveal the constraining or amplifying effects of language knowledge on the effect of strategy use ability across students of different levels of language knowledge. The answer to the second question may help to identify possible reasons leading to different effects of strategy use ability across students of different language knowledge.

### Participants

The study included 1491 nursing students from eight medical colleges in China. All students were aged from 17 to 23 years, with a mean age of 19.7 ( $SD = 1.29$ ). An overwhelming majority of 1453 students (97.5% of the total size) were females and



**Figure 1.** Hypothetical patterns of multi-layered moderation.

only 38 (2.5%) were males. At the time of data collection, all participants had six years' experience of studying English in middle school, one year's experience of studying general English and two months' experience of studying in a nursing English course in college.

## Measures

**Strategy use ability.** Strategy use ability was measured with the Strategy Use Ability Scale (SUAS) (see Table 1). The scale was adapted from previous models of strategy use (Phakiti, 2008a, 2008b; Purpura, 1999) and emphasized the measure of strategy use efficiency (i.e., using the strategy right). The scale had six points, with 0 representing not efficient to 5 representing most efficient. Moreover, the 0 point could also be endorsed when the strategy was not used at all. The scale consisted of 38 items: 15 for metacognitive strategies and 23 for cognitive strategies (see Table 1 for more information). Drawing on Phakiti (2008a, 2008b), the metacognitive strategies subscale were in three categories: planning (e.g., "I considered essential steps needed to complete the reading test"), monitoring (e.g., "I knew when I lost concentration while completing this test"), and evaluating (e.g., "I double-checked my reading comprehension or performance"). The cognitive strategies subscale were also in three categories: memorizing (e.g., "I made notes during the reading"), retrieving (e.g., "I related the information from the reading or tasks to my prior knowledge or experience"), and comprehending (e.g., "I read to see what all or most sentences were in common"). The measurement validity of the SUAS was evaluated using multidimensional item response theory (MIRT) (see Cai, 2013) and the MIRT results showed that a general strategy use ability factor with six domain-specific factors (i.e., planning, monitoring, evaluating, comprehending, retrieving, and memorizing) best represented the SUAS data. The questionnaire was administered in Chinese and the full questionnaire was available in Cai (2013).

**Language knowledge.** Language knowledge (in particular, vocabulary and grammar) has been well documented as a key predictor of reading and L2 proficiency (Jeon

**Table 1.** The Strategy Use Ability Scale (SUAS).

Strategies		Item code	Number of items
Metacognitive strategies	Planning	SU1, SU2, SU3, SU4, SU5	5
	Monitoring	SU6, SU7, SU8, SU9, SU10	5
	Evaluating	SU11, SU12, SU13, SU14, SU15	5
	Subtotal		15
Cognitive strategies	Comprehending	SU16, SU17, SU18, SU19, SU20, SU21, SU22, SU23, SU24, SU25, SU26	11
	Memory	SU27, SU28, SU29	3
	Retrieving	SU30, SU31, SU32, SU33, SU34, SU34, SU35, SU36, SU37, SU38	9
	Subtotal		23
Total			38

& Yamashita, 2014). For this reason, we used a Grammar Knowledge Test (GKT) to measure language knowledge. The GKT used retired items from the Public English Test System-Level Two (PETS-2; National Education Examination Authority, 2007). The PETS-2 in language level corresponds to A2 or B1 of the Common European Framework of Reference for Languages (Council of Europe, 2011). The GKT had 15 discrete sentences, each with a gap to be filled in by selecting the best answer from four options. The items measured students' knowledge of form (nine items) and meanings (six items) at the lexical and syntactical levels (Purpura, 2004).

According to Purpura (2004, 2017), grammatical form refers to linguistic features (phonological/graphological, lexical, morphosyntactic and cohesive features, and so forth) at lexical, sub-sentential, sentential, and supra-sentential levels, whereas grammatical meaning refers to the literal and intended meaning carried by those grammatical forms. Two sample items, the first measuring grammatical form and the second measuring grammatical meaning, are presented here:

*Sample Item 1:* She would rather stay at home than \_\_\_\_\_ with John.

[A] go [B] went [C] going [D] to go

*Sample Item 2:* \_\_\_\_\_ you have finished your work, you are free to do what you like.

[A] Now that [B] Ever since [C] For now [D] By now

The first item was coded as a grammatical form item, as the objective was to test whether students have mastered the correct form of "go" following the expression "would rather". The second sample was designed to test students' ability to appropriately respond using different types of information (meaning) provided to engage in smooth communication. The coding was first done by the first author and further verified by a scholar in assessing



**Table 2.** The Grammar Knowledge Test (GKT).

Components	Elements	Items	M	SD	Number of items	Subtotal
Form	Lexical form	GT3	.39	.49	3	9
		GT12	.27	.45		
		GT14	.24	.43		
	Morphosyntactic form	GT5	.61	.49	4	
		GT7	.35	.48		
		GT11	.36	.48		
		GT15	.33	.47		
	Cohesive form	GT8	.27	.44	2	
		GT13	.42	.49		
Meaning	Lexical meaning	GT6	.68	.47	4	6
		GT9	.46	.50		
		GT10	.38	.49		
	Cohesive meaning	GT2	.78	.41	1	
	Interactional meaning	GT1	.81	.39	2	
		GT4	.72	.45		
Total						15

grammar. The measurement validity of the GKT had been examined using Multidimensional Item Response Theory (see Cai, 2014). As a result, a general factor (i.e., lexico-grammatical knowledge) with two domain-specific factors (i.e., lexico-grammatical form and lexico-grammatical meaning) was found to represent best the GKT data. Descriptive statistics (means and standardized deviations) of the 15 items are presented in Table 2.

*Nursing English (L2) reading ability.* The Nursing English Reading Test (NERT) was used to measure L2 reading ability. The NERT used retired items from the Medical English Test System Level Two (METS-2; METS, 2007). The NERT contained four reading passages, each addressing one of the following topics: gynecology nursing, pediatrics nursing, emergency nursing, and internal medical nursing. Each text had 190 to 300 words, and was accompanied by five multiple-choice questions (see Table 3). The measurement validity of the NERT was examined using multidimensional item response theory (see Cai & Kunnan, 2018). A general nursing English reading ability factor with four domain-specific factors (i.e., gynecology nursing knowledge, pediatrics nursing knowledge, emergency nursing knowledge, and internal medical nursing knowledge) was found to represent best the NERT data. The general factor represented common features shared by all NERT items and each domain-specific factor represented features exclusively shared only by the items/questions of each passage.

### Data collection

Before data collection, ethics approval was obtained from the author's then-host university. Local school permits and participant agreement signatures were obtained from

**Table 3.** The Nursing English Reading Test (NERT).

Text	Topic	Item	Subtotal
Text 1	Dystocia	NR1, NR2, NR3*, NR4, NR5	5
Text 2	Baby Massage	NR6, NR7, NR8*, NR9, NR10	5
Text 3	Emergency	NR11, NR12*, NR13*, NR14, NR15*	5
Text 4	Migraine	NR16*, NR17, NR18, NR19, NR20*	5
Total			20

Note: Items with the star symbol (\*) represent reading for implicit meanings; items without the (\*) represent reading for explicit meanings. NR1 to NR20 = NERT Item 1 to Item 20.

school administrators and students. During data collection, participants were informed of the purpose, background, and general steps to follow. They were first asked to respond to the GKT and NERT on the answer sheets within a time limit of 50 minutes. Immediately after the tests, students were asked to respond to the Strategy Use Ability Survey (SUAS). There was no time limit for the survey response, but no student spent more than 30 minutes on the survey.

### Statistical analyses

The current study used MIRT-based scores obtained through calibrating students' responses to the SUAS, LGKT, and NERT. MIRT is a psychometric model that assumes an individual response to a scale item that is determined by his or her multiple traits ( $\theta$ s) and by multiple item characteristics such as item difficulty and item discrimination. To validate each of the three measures, the authors assumed one general ability underlying all items (e.g., a lexico-grammatical knowledge factor underlying all GKT items) and several uncorrelated domain-specific factors underlying item subsets (e.g., a lexico-grammatical form factor and a lexico-grammatical meaning factor). This structure is known as a bifactor-MIRT (Gibbons et al., 2007). Bifactor-MIRT modeling with each scale involved three major steps: (a) assessing the significance of 1 (general factor) to  $n$  (domain-specific factors) factors; (b) determining the appropriate number of factors; and, accordingly, (c) generating 1 plus  $N$  number of latent scores using the determined bifactor-MIRT structure. According to the MIRT literature, MIRT scores by themselves are meaningless and need to be transformed (Reckase, 2009). Following Reckase, we derived a set of composite scores for each scale by pooling the weighted general factor score and the weighted domain-specific factor scores (thereby,  $n$  composite scores for each scale). One merit of using MIRT-based scores over using the sum-up of raw scores is the reduction of bias owing to item characteristics such as item difficulty and item discrimination (van der Linden & Hambleton, 1997).

As details of the MIRT modeling have been reported elsewhere, this paper only reports analyses directly relevant to the current study. These include the following: (1) descriptive statistics and reliability analyses of MIRT-based scores; (2) confirmatory factor analysis (CFA); (3) multi-layered latent moderation analysis (MLMA); and (4) descriptive statistics of strategy use ability, reported by students of different types, as specified by the MLMA. Descriptive and reliability analyses were run on SPSS Version

20 (IBM Corporation, 2011) and the MLMA was conducted using Mplus 7.4 (Muthén & Muthén, 1998–2018) with the MLR estimator.

Five indices were used for simple SEM model evaluation: the comparative fit index (CFI), the Tucker–Lewis index (TLI), the chi-square ( $\chi^2$ ) statistic, the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Following Mueller and Hancock (2010), a model was considered to have good fit if its CFI and TLI were not smaller than .95 and RMSEA and SRMR values not larger than .05. To evaluate a MLMA model, the statistic of log-likelihood was used. A decision was made by consulting the chi-square significance of  $-2$  times the log-likelihood difference between the simple model and the complex model ( $\Delta-2LL$ ), with the difference between the numbers of free parameters as the degrees of freedom. If the chi-square was significant, then the moderation was justified (Muthén & Muthén, 1998–2017).

## Results

### *Distribution and reliability statistics*

Table 4 presents the descriptive and reliability statistics for the GKT, NERT and SUAS. As these MIRT-based scores were shared features of standardized scores, all mean statistics are near 0 and all standardized deviations are near 1. The means of the SUAS indicators were slightly higher than the means of the GKT and NERT composite indicators (around .10). This was probably because of the larger scale of the SUAS (i.e., six endorsement points: 0 to 5) versus the smaller scale of the GKT and NERT (i.e., two endorsement points: 0 and 1). The skewness and kurtosis values were all within the limits of  $\pm 2$  and suggested reasonably normal distributions.

Cronbach's alphas provided information regarding the consistency of the composite indicators in representing their intended constructs. The GKT composite indicators produced a reliability estimate of 0.75. Given the small number of items ( $j = 2$ ), this value was still considered satisfactory. Cronbach's alphas for the NERT and SUAS were 0.93 and 0.95, respectively, both suggesting high reliability.

### *Model fit results*

The CFA with the full measurement model (combining strategy use ability, language knowledge and nursing English reading ability) produced a good model fit:  $\chi^2 (51) = 133.130$ ,  $p < .001$ ;  $RMSEA = 0.03$  (0.03, 0.04);  $SRMR = 0.02$ ;  $TLI = 0.99$ ;  $CFI = 0.99$ . The standardized loadings for SUAS ranged from .83 (retrieving) to .95 (evaluating), the loadings for the GKT ranged from .76 (meaning) to .79 (form), and the loadings for the NERT ranged from .82 (Texts 1 and 4) to .99 (Text 2).

A structural model was then constructed by regressing L2 reading (NERA) and strategy use ability (SUAS) on language knowledge (GKT). This model was then used as the baseline model (Model 1) to test the three hypothetical moderations: the linear moderation (Model 2), the quadratic moderation (Model 3), and the cuboid moderation (Model 4). The results of model fit are shown in Table 5. The fit indices for the baseline model

**Table 4.** Descriptive statistics and reliability of the composite scores.

Scale	Composite indicator	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	Adjusted Cronbach's <i>a</i>
GKT	Meaning	-0.01	0.91	-0.18	-0.40	-
	Form	-0.01	0.92	0.24	-0.29	-
	Overall					0.75
NERT	Text 1	0.02	0.95	0.17	-0.38	0.92
	Text 2	0.02	0.95	0.30	-0.47	0.87
	Text 3	0.01	0.93	0.21	-0.41	0.91
	Text 4	-0.01	0.95	0.19	-0.19	0.92
	Overall					0.93
SUAS	Comprehending	0.13	0.93	0.01	0.56	0.94
	Memorization	0.13	0.91	0.05	0.58	0.94
	Retrieving	0.14	0.92	0.07	0.57	0.93
	Planning	0.13	0.96	0.06	0.24	0.94
	Monitoring	0.12	0.93	0.02	0.49	0.94
	Evaluating	0.13	0.94	-0.10	0.19	0.94
	Overall					0.95

(Model 1) met the criteria for a good-fit model:  $\chi^2(51) = 133.130, p < .001$ , RMSEA = 0.03 (0.03, 0.04), SRMR = 0.02, CFI = 0.99, TLI = 0.99. The  $-2$  times the log-likelihood for the baseline model (with the number of estimated parameters) was 33,587.63 (39). After successively testing Models 2, 3, and 4 (one following another), the  $\Delta G^2$ s (and associated changes in degree of freedom) were 20.48 (1), 0.22 (1), and 5.76 (1), respectively. According to the chi-square distribution table, two  $\Delta G^2$  values were significant: the linear moderation term (Model 2;  $p < .001$ ) and the cuboid moderation term (Model 4;  $p < .016$ ). The  $\Delta G^2$  produced by the quadratic term was found to be non-significant (Model 3;  $p < 6.36$ ). Building on this evidence of model fit change significance, the highest-layer moderation model was determined as the most appropriate model.

### Results of MLMA

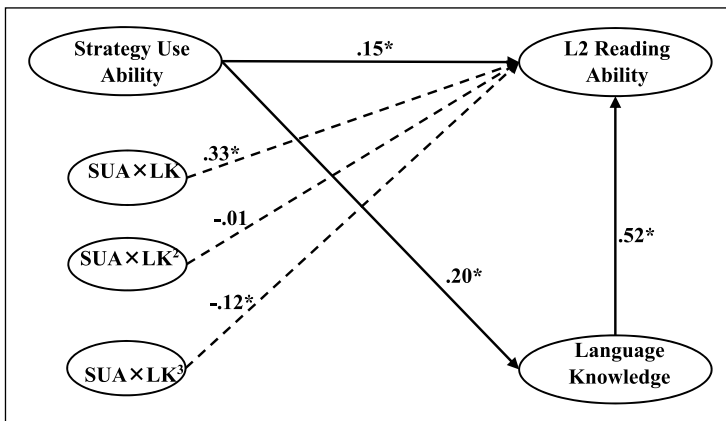
The cuboid moderation model specified the projection of the changing effect of strategy use ability on L2 reading concurrently with the change in language knowledge. The mathematical relationships among L2 reading and strategy use ability, language knowledge, and the interaction between strategy use ability and language knowledge can be expressed as: L2 reading =  $0.48 \times$  language knowledge,  $+(0.16 + 0.25 \times$  language knowledge  $-0.05 \times$  language knowledge cubed)  $\times$  strategy use ability  $+ 0.58$ , where all values were standardized estimates. Following Muthén (2012), the moderation effect of language knowledge on strategy use ability can be expressed as  $0.16 + 0.25 \times$  language knowledge  $-0.05 \times$  language knowledge cubed. The diagram of the resulting structural model is shown in Figure 2.

Plotting language knowledge on the  $x$ -axis and strategy use effect on the  $y$ -axis, this moderation equation produced a projection with two curves (see Figure 3). As shown,

**Table 5.** Model fit results.

Model	AIC	BIC	-2LL	df	$\Delta$ -2LL	$\Delta$ df	p
1	33665.63	33872.61	33587.63	39	-	-	-
2	33647.15	33859.44	33567.15	40	20.48	1	0.000
3	33648.93	33866.53	33566.93	41	0.22	1	0.636
4	33645.17	33868.08	33561.17	42	5.76	1	0.016

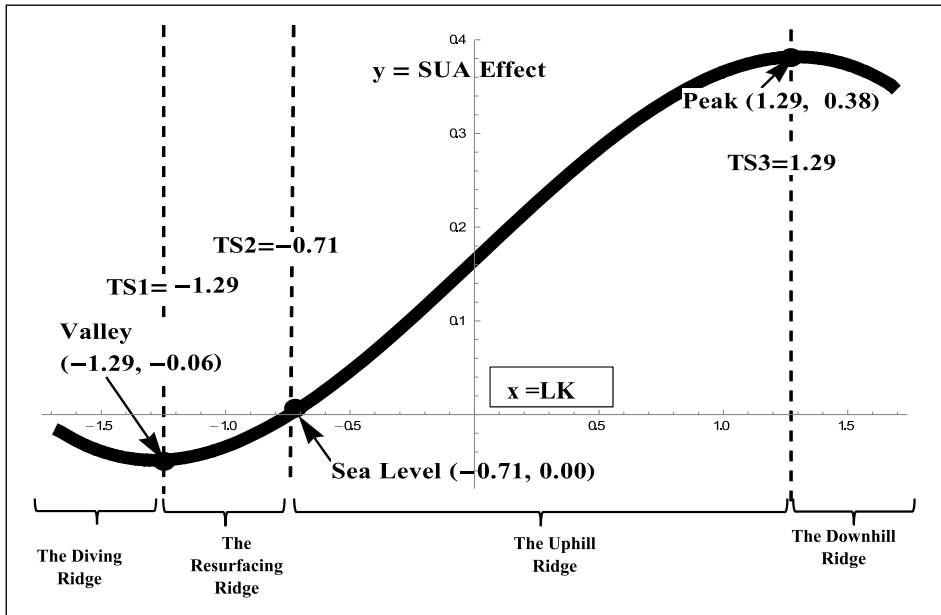
Note: Model 1: The baseline model (main effect model: language knowledge + strategy use ability); Model 2: Layer1 (linear) moderation (language knowledge + strategy use ability + strategy use ability  $\times$  L2 proficiency); Model 3: Layer2 (quadratic) moderation (language knowledge + strategy use ability + strategy use ability  $\times$  language knowledge + strategy use ability  $\times$  language knowledge squared); Model 4: Layer3 (cubed) moderation (language knowledge + strategy use ability + strategy use ability  $\times$  language knowledge + strategy use ability  $\times$  language knowledge squared + strategy use ability  $\times$  language knowledge cubed). In addition, it should be noted that Model 1 was embedded in Model 2, Model 2 was embedded in Model 3, and Model 3 was embedded in Model 4.



**Figure 2.** Diagram for the MLMA (unstandardized estimates).

Note: \* $p < .05$ . SUA  $\times$  LK = product term of strategy use ability with language knowledge (first layer moderation); SUA  $\times$  LK<sup>2</sup> = product term of strategy use ability with language knowledge squared (second layer moderation); SUA  $\times$  LK<sup>3</sup> = product term of strategy use ability with language knowledge squared (third layer moderation).

the curve featured three critical points: the lowest point ( $x = -1.29, y = 0$ ), the intersect point with  $x$ -axis ( $x = -0.71, y = 0$ ), and the highest point ( $x = 1.29, y = 0.38$ ). For the sake of easy communication, these points were labeled metaphorically as the valley, the sea level, and the peak, respectively; and their corresponding  $x$ -axis locations were labeled as the first through the third language threshold. At these three points (or language thresholds), the whole curve was divided into four continuous sections, each with an upward or downward motion direction. Similarly, the four connections were labeled from left to right as the diving ridge, the resurfacing ridge, the uphill ridge, and the downhill ridge. As illustrated by the curve, the whole sample was classified into four different groups: the divers (students in the diving ridge), the resurfacers (students in the



**Figure 3.** The island ridge curve (IRC) illustrating the fluctuating effect of strategy use ability on L2 reading with the increase of language knowledge.

Note: x-axis = LK = language knowledge; y-axis = effect of strategy use ability on L2 reading ability; TS1 = the first language threshold ( $\theta_1 = -1.29$ ); TS2 = the second language threshold ( $\theta_2 = -0.71$ ); and TS3 = the third language threshold ( $\theta_3 = 1.29$ ).

resurfacing ridge), the uphillers (students in the uphill ridge), and the downhillers (students in the descending ridge). In this way, the fluctuating effect of strategy use ability effect on L2 reading with language knowledge can be effectively visualized.

### *Difference in strategy use across students of different L2 proficiency levels*

The difference in students' strategy use of different language knowledge can be examined by studying the means of the six types of strategy distributed across the ridges (see Table 6). For the divers (17 students), the smallest two means were comprehension (-0.29) and planning (-0.26), and the largest two means were memorizing (-0.11) and retrieving (-0.12). Together, this mean distribution produced an overall mean of -0.18 and a standardized deviation of 0.08. As for the resurfacers (169 students), the smallest mean (-0.22) was planning and memorizing, and the largest mean (-0.17) was monitoring and evaluating. The overall mean for the resurfacers was -0.20 with a standardized deviation of 0.02. For the uphillers or the largest group (1275 students), the smallest mean was retrieving (0.14) and the largest was evaluating (0.18). The overall mean for the uphillers was 0.17, with a standardized deviation of 0.01. For the downhillers (30 students), the smallest mean was comprehending (0.80) and

**Table 6.** Means of strategy use efficiency by different strategy users.

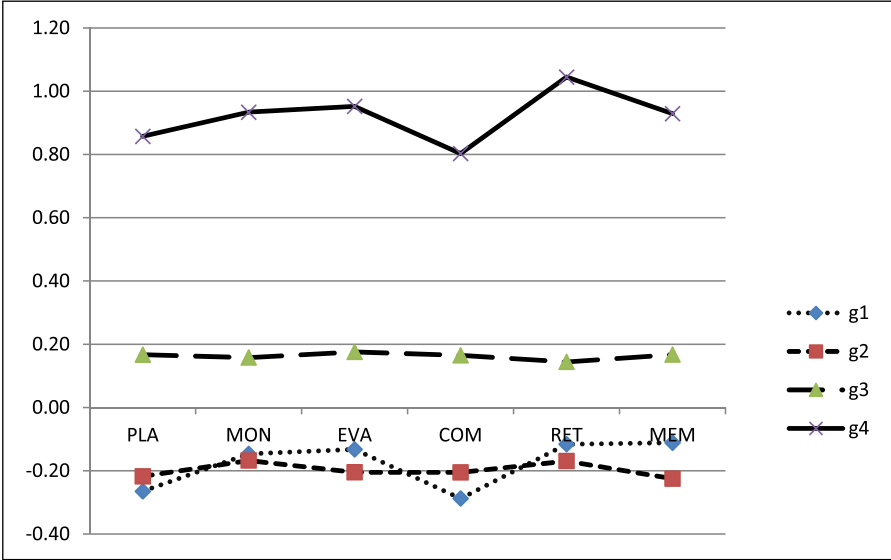
Strategy	Group 1: Divers ( <i>n</i> = 17/1.1%)	Group 2: Resurfacers ( <i>n</i> = 169/11.3%)		Group 3: Uphillers ( <i>n</i> = 1275/85.5%)		Group 4: Downhillers ( <i>n</i> = 30/2%)	
	<i>M</i>	<i>M</i>	Effect size (vs Group 1)	<i>M</i>	Effect size (vs Group 2)	<i>M</i>	Effect size (vs Group 3)
Planning	-.26	-.22	.03	.17	.43	.86	.40
Monitoring	-.15	-.17	.02	.16	.36	.93	.40
Evaluating	-.13	-.20	.07	.18	.42	.95	.40
Comprehending	-.29	-.20	.07	.17	.37	.80	.77
Retrieving	-.12	-.17	.04	.14	.34	1.04	1.00
Memorizing	-.11	-.22	.10	.17	.40	.93	.95
Overall <i>M</i>	-.18	-.20	.06	.17	.39	.92	.65
<i>SD</i>	.08	.02	.03	.01	.04	.08	.29

the largest was retrieving (1.04). The overall mean for this group was 0.92, with a standardized deviation of 0.08.

The effect sizes (Cohen's *d*s) of the mean difference between each two neighboring groups were shown in the right column of Table 6. Overall, the means of the four groups kept ascending all the way from as low as -0.18 with the divers to as high as 0.92 with the downhillers, except for a stumbling of -0.20 with the resurfacers. With respect to the differences in individual means between the divers and the resurfacers, the effect sizes were all trivial except for memorizing ( $d = .10$ ). This information suggested a small but meaningful decrease in using memorizing strategies among the resurfacers when compared with the divers. As for the differences in individual strategy means between the resurfacers and the uphillers, the effect sizes ranged from 0.34 (retrieving) to 0.43 (planning), indicating a medium increase in all strategies. Finally, the individual mean differences between the uphillers and the downhillers yielded effect sizes ranging from 0.40 (with planning, monitoring, and evaluating all being metacognitive strategies) to 1.00 (with retrieving). An illustrative presentation of these summative features is provided in Figure 4.

## Discussion

This study examined the fluctuating effect of strategy use ability on L2 (nursing English) reading ability with language knowledge. Drawing on the MLMA, the current study tested three interaction patterns (i.e., a linear, a quadratic, and a cuboid) by which language knowledge moderated the effect of strategy use ability on L2 reading ability. As a result, the cuboid moderation that took the metaphorical shape of an island ridge curve (IRC) emerged as the most plausible model. The IRC consisted of four continuous connections (i.e., the diving ridge, the resurfacing ridge, the uphill ridge and the downhill ridge), with one joining another at three critical moments (with *x*-, *y*-axis values): the valley (-1.29, -0.06), the sea level (-0.71, 0.00), and the peak (1.29, 0.38), respectively (see Figure 3 for a graphic illustration and Figure 5 for a metaphorical illustration).

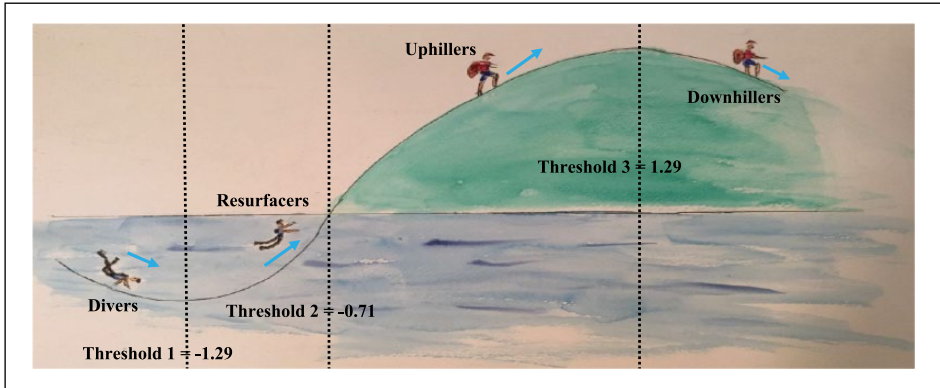


**Figure 4.** Profiles of strategy use within and across the four groups. Note: g1 to g4 = divers, resurfacers, uphillers, and downhillers; PLA = planning; MON = monitoring; EVA = evaluating; COM = comprehending; RET = retrieving; MEM = memorizing. y-axis = strategy use efficiency (MIRT-based scores).

Students within each of these four connections were labeled as the divers, the resurfacers, the uphillers and the downhillers, with each label self-speaking their features of strategy use. Overall, the IRC suggested that as students’ language knowledge increased the effect of strategy use ability on L2 reading fluctuated in a down-up-down pattern.

The diving ridge was located at the lowest continuum of language knowledge and contained a small group of 17 students (for the convenience of communication, labeled as divers since after). Referring to their raw scores in the GKT, they all had the score of 1 out of a maximum of 15. The magnitude of strategy use ability effect on L2 reading for the divers was below the sea level (representing negative magnitude). The continuously downward motion of the projection within this ridge indicated that any increase in language knowledge provided no help in rendering a positive effect of strategy use on comprehension, but instead rendered the situation even worse. This deteriorating situation is mostly because the divers’ L2 proficiency was extremely low (below -1.29 standardized units). According to Perfetti and Hart (2002), students with extremely low language knowledge are usually confronted with much miscoded information. The accumulation of the problem could drive students to a somewhat desperate situation while reading. This desperation again would force students to turn to strategies that were more accessible to them but not closely relevant to the task (Vann & Abraham, 1990). This interpretation can be partly confirmed by examining the diver’s lowest rating on *comprehending*, which is a strategy type regarded as more directly relevant to L2 reading (Chou, 2013; Purpura, 1999). Another explanation is the divers’ inability to master all types of strategies or to put these strategies together into a useful chain of strategies (Green & Oxford,





**Figure 5.** A metaphoric illustration of the island ridge curve (IRC).

1995). This explanation could find support from the relatively large standardized deviation of different strategies within the divers, as compared with those of the uphillers and downhillers (see Table 6). Moreover, strategy use requires additional response time and attention. Therefore, even if some random use of strategies might aid comprehension at the local level, the use occupied the total response time and shifted students' attention away from more appropriate strategies (i.e., *comprehending*) that could aid comprehension at the global level (Walczyk, 2000). As a result, when reading under time pressure, the negative effect became greater (Walczyk, 2000).

The resurfacing ridge revealed the changing effect of strategy use ability on L2 reading with language knowledge between the first and second language thresholds (i.e., between  $-1.29$  and  $-0.71$  standard units, corresponding to 1–3 points, respectively, on the raw GKT scores). For these resurfacers ( $n = 125$ ), language knowledge was still extremely low such that strategy use was still unable to produce a positive effect. However, the chaotic situation gradually eased with the continuous increase in language knowledge. A plausible reason for this upward motion is that, starting from the first language threshold, additionally enhanced language knowledge made the ratio of correct coding more available (Perfetti & Stafura, 2014). The increased availability thus continuously eased resurfacers' desperation, which again enabled them to reduce the activation of certain strategies that were of little use. This reduction, again, allowed students to focus on strategies of greater usefulness (e.g., *planning* and *comprehending*). As language knowledge continued to increase, this positively oriented effect accumulated and eventually "floated up" at the second language threshold. The vivid floating procedure is partly consistent with Clarke's (1980) hypothesis that strategy use is not able to compensate for a lack of language knowledge if language knowledge is extremely low. The added information from the current study is the identification of the exact range that language knowledge can be described as extremely low (i.e., at or below the second language threshold of  $-0.71$ ). Another piece of additional information is the effect size of the compensation for students with extremely low language knowledge. What Clarke's (1980) compensation suggested was indeed a "vacuum" in which strategy use exerted no

effect, either positively or negatively. Our findings of the negative strategy use effect with the divers and resurfacers, however, evidenced the potential of an alternative phenomenon, one that operated with negative effect and varied in a “U”-shaped pattern (i.e., an initially diving motion followed by a switch to a floating-up motion).

The uphill ridge reflected the variation of strategy use ability effect on L2 reading with language knowledge for the uphillers ( $n = 1277$ ). Starting from the second threshold ( $-0.71$ ), the beneficial aspect of strategy use ability was gradually released, moved its way up with the increase of language knowledge, and reached its maximum at the peak of the third language threshold (1.29, corresponding to 11 points on the raw GKT scores). In this sense, the second language threshold corresponds to the lower threshold claimed to short-circuit the transfer of strategy use to L2 reading (Alderson, 2005; Clarke, 1980). The identification of the second threshold supports the critical lexical basis claimed by Perfetti and Hart (2002). According to them, a high enough language knowledge allows students to obtain more information from the written texts by relying on automatized use of language knowledge. This added message again becomes the resources (including time and attention) for the readers to select more relevant strategies (Schoonen, Hulstijn, & Bossers, 1998), or to deploy these strategies more appropriately to accomplish reading (Green & Oxford, 1995; Walczyk, 2000). Alternatively, for the uphillers, their language knowledge has reached such a level that it alone can directly enhance the ability to use strategies (Alderson, 2005). This interpretation can be further supported by studying the means of different strategies across the four groups:  $-0.18$  for the divers,  $-0.20$  for the resurfacers,  $0.17$  for the uphillers, and  $0.92$  for the downhillers. It is interesting to note that the negative/positive values of strategy use efficiency were consistent with the negative/positive values of the actual effect of strategy use on L2 reading.

The downhill ridge illustrated the variation of strategy use effect with language knowledge above the third language threshold of 1.29 (11 points on the GKT raw scores). Like the poorest performers (i.e., the divers), the downhillers also were rather small in number ( $n = 32$ ). For these students, the effect of their strategy use started to descend from its peak at the location of the third language threshold (i.e., 1.29 standardized units) and terminated somewhere at the mid-slope (the highest language knowledge). What this ridge revealed is, when language knowledge becomes extremely high, the potential of beneficial strategy use effect reaches its maximum capacity, such that more use of strategies provides less additional help to comprehension. We hypothesize that if we had been able to include more students of higher language knowledge (i.e., with thetas higher than the highest score in our sample), the descending pattern would be more salient. One might wonder why as a downhiller's language proficiency increases the effect of strategy use decreases, instead of staying as the value of the peak. Checking the means of the six strategy categories in Table 6 (also illustrated in Figure 4), we might get some hints. As shown, the downhillers had relatively smaller values on comprehending and planning, a pattern like the divers, both categories showing descending trend in strategy use ability effect. In contrast, the distribution of the means of the six strategy categories were relatively even with the resurfacers and the uphillers, both showing ascending effect of strategy use ability. It seemed that this uneven distribution of strategy use ability categories played their roles in determining the moving directions of strategy use ability effect with the continuous increase in language knowledge. Our tentative interpretation for these

uneven distributions is that planning and comprehending might be relatively more challenging strategies to master and more energy-consuming to use. Therefore, divers found themselves more unable to use them efficiently, whereas downhillers might be unwilling to engage these more energy-consuming strategies. However, our explanation has yet to be verified in future studies.

This up-down pattern is also like the curvilinear relationship between strategy use and L2 proficiency observed by Hong-Nam and Page (2014). Recall that Hong-Nam and Leavell (2006) explained this curvilinear pattern as the functions of good learners' higher enthusiasm in strategy use. This interpretation, however, seems to be more appropriate for the resurfacers and the uphillers than for the downhillers in the current study. Such an interpretation can find evidence from the means of strategy use (see Table 6). Another interpretation relates to Cohen (2014), according to whom high language knowledge alone can deal with comprehension efficiently and thus makes strategy use less necessary. This alternative interpretation, however, is only minimally supported if we refer to our earlier report regarding the overall strategy use across the four groups (the highest value of 0.92 is by the downhillers). To seek further explanation, we examined the variation (standardized deviation) of the means of the six strategy dimensions across the four groups. We found that the largest value of 0.08 was produced by the downhillers and the divers, as against the values of 0.02 and 0.01 produced by the resurfacers and uphillers, respectively, both moving upwards. It seems the orchestrated use of different strategies plays a certain role in determining the motion direction of strategy use effect. Put another way, the gradual step-down of overall strategy use effect among the top readers could be more of an orchestration issue (i.e., too much retrieving and relatively less comprehending) than a decrease in overall strategy use, as captured in Figure 4. Nevertheless, the evidence from the current study was insufficient to justify the causal relationship between unbalanced strategy use and the decreasing effect of strategy use as a whole among the top readers.

A final point to address is about the divers. At first sight, the diving phenomenon is quite counter-intuitive, and one would very likely wonder if the divers would disappear and the cuboid pattern would reduce to a less complex pattern (i.e., a quadratic or linear pattern), as one of the anonymous reviewers pondered. Our answer may be both yes and no. The answer is yes, if future samples exclude strategy users with L2 proficiency below the first threshold. The answer is no because, even if the divers disappear, this does not necessarily lead to IRC's transformation to a quadratic pattern. This is mostly because a quadratic pattern would have a curve near the middle of the IRC and the cuboid pattern would have two curves near each end of the IRC. Hence, even though the divers may disappear due to sample difference, the pattern will remain cuboid (i.e., the curve will remain near the right end instead of near the middle of the IRC). Similarly, if all sampled students have L2 proficiency below the third threshold, then the downhillers may disappear from a particular study. If only students between the first and third thresholds are sampled, the moderation term then would be reduced to a linear pattern (as there is no data for identifying the curve). Nonetheless, all these three possible patterns remain under the mechanism of the IRC.

## Conclusion

In brief, in this study we found that the effect of strategy use ability on L2 reading ability tended to fluctuate with language knowledge. This fluctuating feature can be summarized in the following five statements. (1) The effect of strategy use ability on language performance fluctuated with the down-up-down pattern with the continuous increase in language proficiency. (2) There were three critical language thresholds that determined the turns of the fluctuation: a low threshold ( $\theta = -1.29$ ), a medium threshold ( $\theta = -0.71$ ), and a high threshold ( $\theta = 1.29$ ). (3) Based on the motion of strategy use effect, strategy users can be classified into four groups: the divers, the resurfacers, the uphillers and the downhillers. (4) The divers and the resurfacers struggled to use strategies effectively and the uphillers and downhillers were beneficiaries of strategy use. (5) The divers and the downhillers were unbalanced strategy users whose strategy use effect was associated with the downwards motion, whereas the resurfacers and the uphillers were balanced strategy users whose strategy use effect was associated with an upwards motion.

Our study has a few limitations. First, our sample size had an extremely unbalanced gender ratio between females and males (97.5% versus 2.5%), an ecological phenomenon of nursing education in China (and perhaps in the world). Second, given the limitation of resources, we were only able to include a list of narrow facets of English language knowledge and nursing knowledge. Future studies may extend these two measures to enhance the representativeness of measurement validity. The third limitation is related to our use of nursing English reading as the dependent variable. Most ideally, we should have included content knowledge as a predictor to let the readers see how multicollinearity between content knowledge and other variables change the pattern of LSP reading. However, given the extreme complexity of the current study, we were only able to include strategy use ability and language knowledge. Therefore, the generalization of the findings should be constrained to the context of Language for Specific Purposes. Finally, as with any other empirical studies, we were not able to recruit students covering a whole range of language proficiency levels. Hence, the exact values of language thresholds should not be taken as fixed and universal, but only as results of our data. If one included other predictor variables in the model, it is possible that the exact locations of the thresholds would vary to a certain extent.

We believe, nonetheless, that our findings could help enhance the field's understanding of the moderation of language knowledge on the effect of strategy use on L2 reading. First, although strategic competence has long been accepted as a component of language ability (Bachman, 1990), this claim has rarely been verified regularly in international language assessment programs. Well-known endeavors in language testing (e.g., Phakiti, 2008a, 2008b; Purpura, 1999) did produce a more comprehensive understanding of strategic competence by accounting for cognitive strategies in the construct, or even the "trait" aspect of strategic competence (Phakiti, 2008a, 2008b). However, a common limitation of their advance is their focus on the "activity" aspect of strategic competence rather than the "ability" aspect of strategic competence. Our study provided a solution by asking test takers to evaluate their efficiency of using these strategies during test performance. To regain the flavor of "competence" as it was originally conceptualized

(Bachman, 1990), more studies need to be conducted in language assessment programs to validate the role of strategy use (especially the “ability” rather than the “frequency” aspect of strategy use) on different language skills (i.e., speaking, listening, writing, and perhaps translation) across L2 learners of different language proficiency.

Second, as indicated by the two underwater ridges, strategy use is not necessarily beneficial for all students’ reading, but can be temporarily harmful as well, for example, when students’ language knowledge is extremely low. This evidence is against the unquestioned belief in the literature that strategy use effect is at most restricted (i.e., no harm). Third, the fluctuating effect of strategy use illustrated by the IRC indicates that the various types of moderation identified in the literature might be merely fragmented excerpts of the same reality captured by observers from their simplified tools. The authors hope that the emergence of the IRC would encourage future researchers to maintain such a relatively holistic stance when exploring this moderation phenomenon, by expanding existing theory and by using other advanced analytical approaches.

Our results should also convey useful information for reading strategy training. It is important for teachers to understand the differentiated role of strategy use for L2 readers of different language knowledge. It is advisable that, before plunging into strategy training, teachers use their intuitive judgment to group students into extremely low-, low-, medium, and high-language proficiency groups. Wherever possible, teachers with quantitative skills might use conventional cluster analysis to put students into different language proficiency groups, if MLMA is too complicated to apply. Strategy training programs may then focus on students of medium language proficiency. For students of low language proficiency, it is advisable for them to spend time improving their general language proficiency (e.g., knowledge of lexical and syntactical form and meaning) before they invest their time in enhancing their strategy use ability.

Finally, our study was a synergy of theory and methodology. At the time of the current study, no existing analytical approach was found in the literature that could be directly applied to tap into the nonlinear moderation pattern(s) between language knowledge and strategy use. Driven by this substantial need, we took great pains to create the MLMA and test its validity with real data. This theory-driven way of enquiry thus provided a greater opportunity for the authors to see the truth which would otherwise have been veiled if a less rigorous analytical approach such as multigroup regression analysis were used. In terms of theory, interaction has been conceptualized as an essential phenomenon in language testing literature. For instance, Bachman and Palmer (1996, 2010) emphasized that a key issue for CLA is the interaction between language knowledge and strategic competence, and between these CLA components and other construct-irrelevant elements (e.g., test takers’ individual characteristics such as gender, test methods, and so forth). Surprisingly, few empirical studies have systematically examined these possible interactions. To ensure the quality of language tests, more studies are needed to investigate such interactions. The MLMA should provide a rigorous tool for this stream of exploration.

### **Acknowledgements**


The authors would also like to extend their sincere thanks to scholars who have generously provided invaluable consultations on the project: Professor Bengt Muthen and Professor Zhonglin

Wen on latent interaction, Professor Mark Reckase and Professor Li Cai on multidimensional item response theory, and Professor Jim Purpura on the coding of the grammar knowledge test. Regardless, all possible errors remain those of the authors. The authors' thanks also go to Dr Iris Lin for helping the authors to draw the lovely IRC metaphoric picture.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work was partly supported by three grants offered to the first author: (1) The Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (Code: TP2018068), (2) the TOEFL Small Grants for Doctoral Research in Second or Foreign Language Assessment, Educational Testing Service, USA, and (3) the Grants for Graduate Students in Psychological and Educational Measurement Programs, Assessment Systems Corporation, USA.

## ORCID iD

Yuyang CAI  <https://orcid.org/0000-0002-0320-4602>

## References

- Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. New York: Continuum.
- Alsamadani, H. A. (2009). The relationship between Saudi EFL college-level students' use of reading strategies and their EFL reading comprehension. (PhD thesis). Ohio University, USA.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Hong Kong: Oxford University Press.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, UK: Oxford University Press.
- Cai, Y. (2013). Validating a scale measuring strategic competence in ESP reading test context: An application of bifactor-MGRM. *The Asian ESP Journal* [Special Edition], 9(2), 26–51. [www.asian-esp-journal.com/journal-index/2013/october-2013/](http://www.asian-esp-journal.com/journal-index/2013/october-2013/)
- Cai, Y. (2014). Comparing two theories of grammatical knowledge assessment: A bifactor-MIRT analysis. *Language Learning in Higher Education*, 4(1), 59–76. <https://doi.org/10.1515/cercles-2014-0005>
- Cai, Y., & Kunnan, A. J. (2018). Examining the inseparability of content knowledge from LSP reading ability: An approach combining bifactor-multidimensional item response theory and structural equation modeling. *Language Assessment Quarterly*, 15(2), 109–129. <https://doi.org/10.1080/15434303.2018.1451532>
- Cai, Y., & Kunnan, A. J. (2019). Detecting the language thresholds of the effect of background knowledge on a Language for Specific Purposes reading performance: A case of the island ridge curve. *Journal of English for Academic Purposes*, 100795. <https://doi.org/10.1016/j.jeap.2019.100795>
- Chou, M.-h. (2013). Strategy use for reading English for General and Specific Academic Purposes in testing and nontesting contexts. *Reading Research Quarterly*, 48(2), 175–197. <https://doi.org/10.1002/rq.42>

- Clarke, M. A. (1980). The short circuit hypothesis of ESL reading—or when language competence interferes with reading performance. *Modern Language Journal*, 64(2), 203–209. <https://doi.org/10.2307/325304>
- Cohen, A. D. (2008). Speaking strategies for independent learning: A focus on pragmatic performance. In S. Hurd & T. Lewis (Eds.), *Language learning strategies in independent settings* (pp. 119–140). Toronto: Multilingual Matters.
- Cohen, A. D. (2014). *Strategies in learning and using a second language*. New York: Routledge.
- Council of Europe. (2011). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. [https://www.coe.int/t/dg4/linguistic/source/framework\\_en.pdf](https://www.coe.int/t/dg4/linguistic/source/framework_en.pdf)
- de Larios, J. R., Manchón, R., Murphy, L., & Marín, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing*, 17(1), 30–47. <https://doi.org/10.1016/j.jslw.2007.08.005>
- Dörnyei, Z., & Ryan, S. (2015). Learning strategies and self-regulation. In Z. Dörnyei & S. Ryan (Eds.), *The psychology of the language learner revisited*. New York and London: Routledge.
- Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge, UK: Cambridge University Press.
- Gibbons, R. D., Bock, R. D., Hedeker, D. R., Weiss, D. J., Segawa, E., Bhaumik, D. K., . . . Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, 31(1), 4–19. <https://doi.org/10.1177/0146621606289485>
- Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading* (2nd ed.). London: Pearson.
- Green, J. M., & Oxford, R. (1995). A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly*, 29(2), 261–297. <https://doi.org/10.2307/3587625>
- Griffiths, C. (2013). *The strategy factor in successful language learning*. Bristol, UK: Multilingual Matters.
- Griffiths, C., & Inceçay, G. (2016). New directions in language learning strategy research: Engaging with the complexity of strategy use. In *New directions in language learning psychology* (pp. 25–38). New York: Springer.
- Gürses, M. Ö., & Bouvet, E. (2016). Investigating reading comprehension and learning styles in relation to reading strategies in L2. *Reading in a Foreign Language*, 28(1), 20–42. <https://eric.ed.gov/?id=EJ1098658>
- Hong-Nam, K., & Leavell, A. G. (2006). Language learning strategy use of ESL students in an intensive English learning context. *System*, 34(3), 399–415. <https://doi.org/10.1016/j.system.2006.02.002>
- Hong-Nam, K., & Page, L. (2014). Investigating metacognitive awareness and reading strategy use of EFL Korean university students. *Reading Psychology*, 35(3), 195–220. <https://doi.org/10.1080/02702711.2012.675418>
- IBM Corporation. (2011). IBM SPSS Statistics for Windows, Version 20.0 [computer program]. Armonk, NY: IBM Corporation.
- Jeon, E. H., & Yamashita, J. (2014). L2 Reading Comprehension and its correlates: A meta-analysis. *Language Learning*, 64(1), 160–212. <https://doi.org/10.1111/lang.12034>
- Jiménez, R. T., García, G. E., & Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly*, 31(1), 90–112. <https://doi.org/10.1598/RRQ.31.1.5>
- Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65(4), 457–474. <https://doi.org/10.1007/BF02296338>
- Magogwe, J. M. (2013). Metacognitive awareness of reading strategies of University of Botswana English as Second Language students of different academic reading proficiencies. *Reading & Writing*, 4(1), 1–8. <https://doi.org/10.4102/rw.v4i1.29>

- METS. (2007). *Medical English Test System Level Two (for Nurses)*. Shanghai: Higher Education Press.
- Mokhtari, K., & Sheorey, R. (2002). Measuring ESL students' awareness of reading strategies. *Journal of Developmental Education, 25*(3), 2–11.
- Mueller, R. O., & Hancock, G. R. (2010). Structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 373–383). New York: Routledge.
- Muthén, B. Q. (2012). Latent variable interactions. <http://statmodel2.com/download/Latent%20variable%20interactions.pdf>
- Muthén, L. K., & Muthén, B. Q. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. Q. (1998–2018). *Mplus 8.2* [Computer software]. Los Angeles, CA: Muthén & Muthén.
- National Education Examination Authority. (2007). *Public English Test System-Level Two*. Beijing: Higher Education Press.
- O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge, UK: Cambridge University Press.
- Oxford, R. (1990). *Language learning strategies: What every teacher should know*. Boston, MA: Heinle & Heinle.
- Oxford, R. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed.). New York: Routledge.
- Oxford, R., & Amerstorfer, C. M. (2018). *Language learning strategies and individual learner characteristics: Situating strategy use in diverse contexts*. London: Bloomsbury.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. T. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (Vol. 11, pp. 67–86). Amsterdam and Philadelphia, PA: John Benjamins.
- Perfetti, C. A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22–37. <https://doi.org/10.1080/10888438.2013.827687>
- Phakiti, A. (2008a). Construct validation of Bachman and Palmer's (1996) strategic competence model over time in EFL reading tests. *Language Testing, 25*(2), 237–272. <https://doi.org/10.1177/0265532207086783>
- Phakiti, A. (2008b). Strategic competence as a fourth-order factor model: A structural equation modeling approach. *Language Assessment Quarterly, 5*(1), 20–42. <https://doi.org/10.1080/15434300701533596>
- 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. (2014). Cognition and language assessment. In A. J. Kunna (Ed.), *The companion to language assessment* (pp. 1452–1476). Oxford, UK: Wiley-Blackwell.
- Purpura, J. E. (2017). Assessing meaning. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., pp. 33–61). New York: Springer.
- Razi, O., & Grenfell, M. J. (2012). The impact of linguistic knowledge on learner strategy deployment. *Procedia-Social and Behavioral Sciences, 47*, 818–822. <https://doi.org/10.1016/j.sbspro.2012.06.741>
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
- Rubin, J. (1975). What the “good language learner” can teach us. *TESOL Quarterly, 9*(1), 41–51. <https://doi.org/10.2307/3586011>
- Rubin, J. (2005). The expert language learner: A review of good language learner studies and learner strategies. In K. Johnson (Ed.), *Expertise in second language learning and teaching* (pp. 35–63). New York: Palgrave Macmillan.



- Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8 and 10. *Language Learning*, 48(1), 71–106. <https://doi.org/10.1111/1467-9922.00033>
- Shorey, R., & Mokhtari, K. (2001). Differences in the metacognitive awareness of reading strategies among native and non-native readers. *System*, 29(4), 431–449. [https://doi.org/10.1016/S0346-251X\(01\)00039-2](https://doi.org/10.1016/S0346-251X(01)00039-2)
- Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.
- Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, 27(1), 78–102. <https://doi.org/10.1093/applin/ami046>
- Van der Linen, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. New York: Springer.
- Vandergrift, L. (2008). Learning strategies for listening comprehension. In S. Hurd & T. Lewis (Eds.), *Language learning strategies in independent settings* (pp. 84–102). Toronto: Multilingual Matters.
- Vann, R. J., & Abraham, R. G. (1990). Strategies of unsuccessful language learners. *TESOL Quarterly*, 24(2), 177–198. <https://doi.org/10.2307/3586898>
- Walczyk, J. J. (2000). The interplay between automatic and control processes in reading. *Reading Research Quarterly*, 35(4), 554–566. <https://doi.org/10.1598/RRQ.35.4.7>