Educational Research and Development Journal
Fall 2020, Vol. 23, No. 2, pp. 22–38

Investigating the Theoretical Structure of the Survey of Reading Strategies

Tiejun Zhang
Jiali Zheng
College of Education, University of South Carolina

Abstract

The purpose of this study is to investigate the theoretical structure of the Survey of Reading Strategies (SORS) developed for assessing ESL students' use of metacognitive reading strategies. The data for this study were collected from 1,431 Chinese college students whose second language (L2) was English. Both explanatory factor analysis (EFA) and confirmatory factor analysis (CFA) were conducted to validate the SORS. The EFA failed to recover the three hypothesized subscales but supported a unidimensional construct. The CFA analysis suggested that all alternative factor models fitted the sample data poorly, but comparative fit indices indicated the construct of the SORS was unidimensional. Therefore, the factor structure of the SORS needs to be further explored. Suggestions and recommendations for the use of the survey were provided.

Investigating the Theoretical Structure of the Survey of Reading Strategies

In reading comprehension, many researchers are interested in how readers monitor their reading process and what strategies they tend to use to facilitate their reading comprehension, which is often referred to as metacognitive awareness and perceived use of reading strategies. Studies suggested that metacognition of reading strategies was associated with learner's reading comprehension performance (Flavell, 1979; Paris & Winograd, 1990; Mokhari & Reichard, 2002). Therefore, investigating reading strategies use not only helps researchers distinguish skillful readers from poor readers, but also identifies effective reading strategies that can be used to enhance reading comprehension. Many metacognition scales were developed by researchers to measure the use of reading strategies (Jacobs & Paris, 1987; Miholic, 1994; Pereira-Laird & Deane,1997; Mokhari & Reichard, 2002a, 2002b; Zhang, 2018). For researchers particularly concerned with L2 reading comprehension, the Survey of Reading Strategies (SORS) developed by Mokhari & Reichard (2002b) is one of the most widely used scales (See Victoria, 2012; Hong-Nam & Page, 2014; Yaemtui, 2015; Miller, 2017; etc.). This 30-item survey is comprised of three subscales: global reading strategies, problem-solving strategies, and support strategies. Previous studies have reported high reliability for the scale, but its construct validity, namely, the theoretical three-factor structure, has not been adequately validated. The SORS was adapted from the Metacognitive Awareness of Reading Strategies Inventory (MARSI, Mokhari & Reichard, 2002a) that was designed to measure metacognition among young native English learners. Mokhari and Reichard (2002b) only field-tested the internal consistency of the scale on ESL students. Based on theoretical considerations and factor analyses on MARSI, they claimed that the SORS measures three broad reading strategies: global reading strategies, cognitive strategies, and support strategies. Therefore, the hypothesized structure of the SORS was not validated by factor evidence obtained on a population of L2 learners. This leaves room for concerns over the construct validity of the SORS, as it is a measure developed for L2 learners.

Literature Review

Metacognition and Reading Comprehension

Metacognition, simply put, is the thinking about thinking. It refers to the awareness and understanding of one's cognitive processes. Metacognition in reading often includes metacognitive awareness and metacognitive regulation (Paris & Winograd; 1990; Mokhari & Reichard, 2002a; Van Kraayenoord, 2010). Metacognitive awareness relates to readers' awareness or knowledge of their comprehension processing and strategies that support their comprehension, while metacognition regulation"involves cognitive activities such as planning, checking, evaluating and testing and revising strategies" (van Kraayenoord, 2010, p. 278).

Metacognition plays a vital role in successful reading comprehension. The use of metacognitive strategies is associated with learners' reading competence. Studies indicate that good readers often have better metacognitive awareness and regulation than poor readers. Good readers were more aware of their reading purpose and what they read, and they tended to use more strategies and use them more frequently than poor readers (Jacobs & Paris, 1984; Snow, Burns & Griffin, 1998). In terms of teaching practice, researchers have also revealed that students' metacognitive awareness can be improved through systematic direct instruction (Paris & Winograd, 1990).

Researchers revealed that cognitive activities differ from L2 reading to L1 reading. L2 reading is mainly affected by the reader's L1 transfer and L2 proficiency, which leads to a difference in the use of reading strategies between L1 and L2 reading (Grabe, 2008, Tsai, et al, 2010; Jou, 2015). Sheorey and Mokhtari (2001) compared the use of three categories of strategies between 150 native-English U.S. and 152 ESL college students. They concluded that both skilled native and ESL students reported higher usage of cognitive and metacognitive reading strategies than their less-skilled counterparts, but ESL students all valued support strategies while only skilled readers of native students attached greater importance to them.

Many scales have been developed to measure metacognitive reading strategies (Jacobs & Paris, 1987; Miholic, 1994; Pereira-Laird & Deane, 1997; Mokhari & Reichard, 2002a, 2002b; Taraban, et al, 2004; Zhang, 2018). Here we only focused on the SORS and the MARSI from which the SORS was adapted, because the SORS was the only scale developed for assessing L2 learners' use of metacognitive reading strategies.

The MARSI and the SORS

The MARSI (Mokhari & Reichard, 2002a) was developed to assess adolescent and adult readers' metacognitive awareness and perceived use of reading strategies while reading academic or school-related materials. Factor analytic evidence suggested that the 30-item scale consisted of three subscales: global reading strategies, problem-solving strategies, and support strategies. Later, Mokhari and Reichard (2002b) modified MARSI and adapted it into the SORS so it could be used with adolescent and adult students whose second language is English. The SORS was assumed to enjoy the same three subscales as the MARSI, given the association between them. Consequently, the construct validity evidence for the SORS mainly came from the validation of the MARSI (Mokhari & Reichard, 2002b).

In measuring a latent trait that cannot be directly observed, construct validity determines the extent to which a scale captures the hypothetical construct by a set of indicators. In other words, construct validity provides evidential basis for score interpretation (Messick, 1995). Construct validation procedures often provide convergent and discriminant evidence for score meaning and interpretation. Furthermore, it is part of the theory testing process to determine whether a measure validly represents a hypothetical construct (Smith, 2005).

Mokhari and Reichard (2002a) validated the MARSI in two steps. First, they constructed a 60-item instrument out of an item pool of 100 reading strategies, and it was field-tested with a sample of 825 adolescent students from 10 districts in five middle east states. The EFA analyses suggested a three-factor model of the initial instrument. Half of the items were dropped due to low factor loading, cross-loading, reduced reliability, and duplication. Then, the finalized 30-item scale was again tested with a similar sample of 443 students. According to their findings, the EFA analyses supported the three-factor model which accounted for 29.7% of the total variance. The reliability (Cronbach's alpha) was .89 for the sample. However, it is worth pointing out that cross-loading issues were not well addressed in their three-factor solution. Five items almost equally loaded on two different factors. In addition, CFA was not conducted to verify this hypothesized model. Wu et al. (2012) translated the MARSI into Chinese and utilized it to measure Chinese middle school students' metacognition in reading Chinese. The EFA analysis with a sample of 494 students successfully recovered the theoretical three subscales, but their criteria for dropping items were not well justified. They simply chose to remove the items with loading values lower than .50 and items that were not clustered within the hypothetical three factors. Eventually only 16 items were retained. In their CFA analysis, the implication of high correlations among the three factors (.79 ~ .91) was not thoroughly discussed. To address issues about MARSI's appropriateness for college and adult readers, its association with reading ability, as well as generalizability issues, Mokhtari et al. (2018) revised the MARSI into a 15-item instrument and retained the original three subscales regardless of the revisions. Besides, they only relied on modification indices (MIs) for removing items. MIs indicate the extent to which the model will improve if a particular parameter was added to the model; however, the use of MIs is often criticized because they are more based on statistical considerations rather than theoretical evidence (Bandolas, 2018).

Despite the inadequate investigation of the hypothesized subscales of the MARSI, another concern is to what extent the validity evidence for the MARSI can be used to validate the SORS. As a scale particularly concerning L2 learners, there is a lack of validity evidence directly obtained from a population of L2 learners.

In view of the research gaps, it is of both practical and theoretical importance to examine the underlying structure of the SORS. Factor analytic studies as a method of construct validation should be an ongoing process, and subsequent studies could help to test the construct dimensions formed in previous studies (Mulaik, 2010). This study aims to investigate the underlying structure of the SORS with data collected on L2 English learners in China. EFA was conducted to test whether the theoretical structure of the SORS could be recovered and supported and to explore other alternative explanations. Additionally, CFA was used to compare model fit to the sample data among competitive models.

Method

Instrument

Data for this study were collected through an online survey that consists of two questionnaires, the background questionnaire and the survey of reading strategies. The background questionnaire was developed by the researchers to elicit demographic information about the participants. It consists of 11 items, which covers school name, age, gender, major, years of English studying, time spent on English reading per week, reasons for English reading, attitudes towards English reading, types of reading materials, and self-evaluated English reading level.

The survey of reading strategies (SORS) (Mokhari & Reichard, 2002b) comprises 30 items that are categorized into three subscales: global reading strategies (GLOB), problem-solving strategies (PROB), and support strategies (SUP). GLOB strategies (13 items) refer to intentional, carefully planned techniques such as having a purpose in mind when reading, previewing the text length and organization. PROB strategies (8 items) are the actions and procedures to comprehend the text, such as adjusting reading speed and guessing the meaning of unknown words. SUP strategies (9 items) refer to the use of resources or methods to aid reading, such as looking up a word in the dictionary, taking notes. Each item presents a statement about a particular reading strategy that should be responded on a five-point Likert scale (1 – I never or almost never do this, 2 – I do this only occasionally, 3 – I sometimes do this, 4 – I usually do this, 5 – I always or almost always do this).

Some revisions were made to the SORS to accommodate the characteristics of Chinese participants. For example, phrases like "mother tongue" and "native language" were substituted by the more specific term "Chinese." Additionally, the scale was translated from English into Chinese to avoid potential comprehension difficulty. To ensure the quality and accuracy, the translation work was done by three native Chinese speakers who were proficient in English, including the two authors. They were all doctoral students in education at a southeastern U.S. university. The Chinese version of the scale was further reviewed by a Chinese professor in language and literacy at the same university. The translated survey was compiled on a questionnaire website, and the link was sent to eligible participants in China via WeChat, a mobile app that is widely used by residents in China.

As for scale reliability, Cronbach's alpha or coefficient alpha (Cronbach, 1951) is often reported by researchers to illustrate the internal consistency among items of a scale. The overall Cronbach alpha of the SORS for this study was as high as .94, indicating that all items were highly correlated. The alpha coefficients for the three subscales were as follows: GLOB strategies ($r = .90$), PROB strategies ($r = .91$), and SUP strategies ($r = .92$). These high alpha values suggested items within each subscale were also highly correlated.

Participants and Data

The data for this study were collected from 1,431 Chinese undergraduate and graduate students in 2018. The sample consisted of 1,081 females and 350 males aged between 16 and 43, with a mean age of 20.6. The participants who majored in over ten areas of study were from 60 different colleges and universities across China. Participants' years of experience in academic English reading ranged from one to three years, with a mean of 1.88. The snowballing sampling method (Goodman, 1961) was employed to recruit eligible participants. We reached out to university faculty and staff that we knew and asked them to send the survey to eligible college and graduate students. The translated SORS was administered online, reaching out to eligible participants. Altogether 2,342 questionnaires were received. The data cleaning was mainly based on two procedures: the response time and response pattern. A pilot study with 25 students suggested that a reasonable time to read and complete the survey was about five minutes. Therefore, responses completed within four minutes were removed for validity concerns. Moreover, those observations with a clear response pattern were taken out. A typical case was that some participants selected the same option for all items. According to the above two criteria, 910 observations were deleted, and 1,431 responses were considered valid after data cleaning.

Statistical Methods

Factor analysis is a statistical method used in measurement to develop or verify a theory. As Bandolas (2018) stated: "the purpose of factor analytic techniques is to help us to understand the number and nature of the dimensions, or factors, underlying a set of variables" (p. 301). Two commonly used factor analysis methods are explanatory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is often used to determine how many factors underlie the hypothesized structure, while CFA is often used to test whether our understanding of the factor model of the hypothesized structure is consistent across time or sample. In other words, EFA seeks to generate theory about a construct while CFA focuses on verifying theories about a construct's dimensionality.

According to Bandalos (2018), it is inappropriate to conduct both EFA and CFA using the same dataset because it leads to an overfitting problem. The factor solution obtained from the EFA tends to fit the data better than other alternative models in the CFA analysis if the same sample is used in both procedures. A solution to address this issue is to split the dataset into two random sub-datasets, using one sub-dataset for EFA and the other sub-dataset for CFA. In this study, the 1,431 participants were randomized into two sub-datasets. The EFA was conducted with a sample of 715 respondents, and the CFA was conducted with a sample of 716 respondents.

EFA Procedures

Explanatory factor analysis was performed with 30 items of the SORS using a sample of 715 participants. The ratio of the number of observations to the number of variables is about 1:23, meeting the suggested requirements for factor analysis (Bandolas, 2018).

Factor extraction method. Principal component analysis (PCA) and principal axis factoring (PAF) are two commonly used methods to extract factors. The major difference between PCA and PAF is that PCA analyzes all the variance in the manifest variables (items), including the covariance and unique variance(error), while PAF only analyzes the covariance. Factor analysis assumes that latent variables cause manifest variables to covary; therefore, PAF is preferable to PCA in practice. In this study, the PAF method was employed to extract factors.

Number of Factors to Retain. After choosing the factor extraction method, the next step is to determine the appropriate number of factors to retain. Such techniques include the "K1" criterion (Kaiser, 1960), scree test (Cattell, 1966), parallel analysis (Horn, 1965), etc. The K1 criterion simply retains all the components with eigenvalues greater than one, but it always suffers from an overestimation of the number of factors. The scree plot is a graph presenting the relationship between the eigenvalues and the number of factors. The eigenvalue is the largest for the first factor extracted, then will decrease by each additional factor extracted. Therefore, only the first of a few factors will have large eigenvalues and the latter factors will have small and similar eigenvalues, which results in a sharp bend or break flattening the downtrend of the eigenvalues on the scree plot. The number of factors before the bend is the optimal number to retain. Parallel analysis (Horn, 1965) compares eigenvalues from the actual data to the average eigenvalue obtained from randomly generated datasets. These random datasets are generated based on the actual data. Factors are retained if the actual eigenvalue is greater than the average eigenvalue from the random data set. Studies show that the K1 rule and scree plot tend to overestimate the number of factors, and PA is much more accurate in most situations. Therefore, researchers always use multiple criteria to determine the optimal number of factors to retain and try different solutions. A combination of these three methods was employed to determine the optimal number of factors to extract in this study.

Rotation. The purpose of the rotation is to minimize the complexity of the factor loadings and make the structure easier to interpret. There are orthogonal and oblique rotations. Orthogonal rotations assume factors are uncorrelated, whereas oblique rotations allow factors to be correlated. Oblique rotations are preferred in factor analysis because they could result in the same results using orthogonal rotations if the factors are uncorrelated. Consequently, the oblique rotation was used for explanatory factor analysis throughout the study.

Evaluation Criterion. Distefano and Dombrowski (2006) proposed five criteria to evaluate EFA solutions: First, examining the percentage of variance explained by the overall factor model and by each factor. Second, examining the simple structure, that is, each item should load strongly on only one factor. An item with a loading value above .30 is considered a good indicator of a factor. An item is cross-loading if it loads on more than one factor and all loading values are above .30. Third, examining the absence of specific factors in a solution. Fourth, examining the residual matrix. Large residual terms imply that there are additional factors still to be extracted. Finally, examining the interpretability of the factor solution and its match to theory. Costello and Osborne (2005) argued that "A factor with fewer than three items is generally weak and unstable; 5 or more strongly loading items (.50 or better) are desirable and indicate a solid factor". In this study, we concentrated on the interpretability of the solution and the simple structure because they were the two most important criteria for evaluating EFA solutions.

CFA Procedures

Alternative factor models. For confirmatory factor analysis, a series of alternative models were tested with 715 participants. Based on previous studies and our EFA results, four alternative factor models of the SORS were eventually identified: (a) general factor model; (b) reduced general factor model; (c) three-factor model; (d) reduced three-factor model.

The general factor model was supported by our EFA results, indicating all 30 items measured a unidimensional construct. Regarding the reduced general factor model, Mokhtari et al. (2018) reduced the number of MARSI items from 30 to 15, and only 14 items were retained in the SORS. Therefore, the reduced general factor model only contained 14 items as compared to the general factor model. The three-factor model is the hypothetical structure of the SORS. The three subscales are "Global Reading Strategies" (13 items), "Problem Solving Strategies" (8 items), and "Support Strategies" (9 items). Similarly, the reduced three-factor model retained the same structure with 14 items compared to the three-factor model. The subscales are "Global Reading Strategies" (5 items), "Problem Solving Strategies" (5 items), and "Support Strategies" (4 items).

Estimator. The default maximum likelihood estimation (ML) was used for the CFA analysis in this study. ML estimators are unbiased and efficient when multivariate normal distribution assumption is met. The descriptive statistics showed that these items were approximately normally distributed.

Model Fit Criteria. Researchers rely on model fit information to evaluate the extent to which a factor model fits to a sample dataset and compare different CFA models. The chi-square statistic(x^2) and some other fit indices recommended by Hu and Bentler (1999) were commonly used in CFA analysis. The chi-square statistic(x^2) tests the overall model fit of a model to a sample data. A non-significant statistic ($p > .05$) suggests the model fits the data well. However, this index is sensitive to sample size. It is very likely to get a significant value with a large sample. The comparative fit index (CFI), Tucker-Lewis index (TLI), and the global fit index (GFI) all compare the fit of the tested model to the baseline model, and values over .95 indicate a good model fit of the tested model. The standardized root mean square residual (SRMR) estimates the average squared discrepancy between the model and sample covariance matrices, and values under .08 suggest acceptable model fit. The root mean square error of approximation (RMSEA) measures how closely the model reproduces data patterns and values less than .06 demonstrate a close fit. The 90% confidence interval (CI) around the RMSEA point estimate should contain .05 to indicate the possibility of close fit. As the x^2 statistic was likely to be biased due to the large sample size, we chose to focus on CFI, TLI, GFI, SRMR, RMSEA and its 90% CI to evaluate and compare the CFA models.

Results

EFA Results

The Scree plot and parallel analysis methods were employed to determine the number of factors to retain. Figure 1 presents the scree plot analysis of the SORS data. Only one factor can be extracted by the K1 criterion (eigenvalue larger than 1). The scree plot also suggests one factor (before the "elbow"). The PA method indicated that the optimal number of factors to retain was two. However, the theoretical structure contains three factors. As the optimal number of factors did not converge, solutions with 1 to 4 factors were all examined. Bandolas (2018) suggested that any number of factors supported by either method or theory and prior study should be explored in EFA, and solutions with one more or one less factor than the recommended number of factors should also be considered. The results of the four EFA solutions were carefully weighted and compared based on the aforementioned criteria.

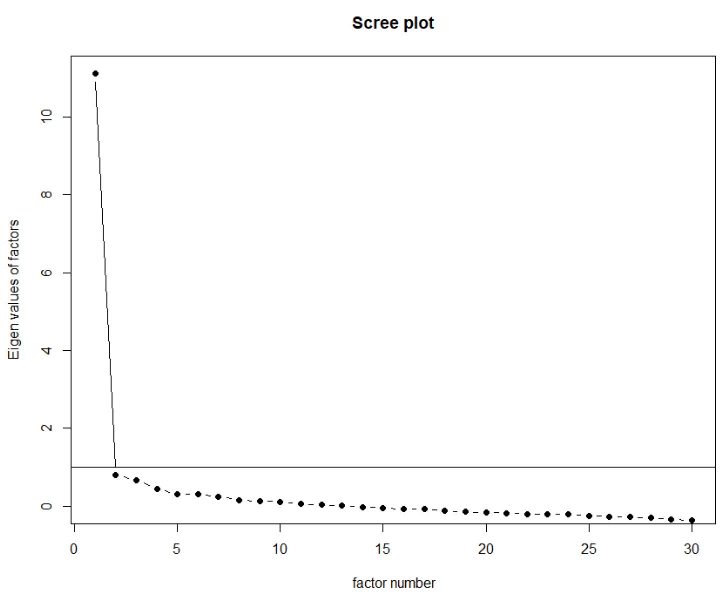

Figure 1. Scree Plot of the SORS

The *one-factor solution* accounted for 37% of the variance (see Table 1). Items were at least moderately loaded on the general factor, with loading values ranging from .38 to .74, indicating these items were good indicators of the general factor. This model was retained to compare with other solutions.

Table 1

Factor Analysis Results for SORS Unidimensional Model (N =715)

Items	Hypothesized Factors*	Factor Loading (Cutoff = .30) F1
1. I have a purpose in mind when I read English text related to my content area.	GLOB	0.53
2. I take notes while reading to help me understand what I read.	SUP	0.46
3. I think about what I know to help me understand what I read.	GLOB	0.64
4. I take an overall view of the text to see what it's about before reading it.	GLOB	0.51
5. When text becomes difficult, I read aloud to help me understand what I read.	SUP	0.39
6. I think about whether the content of the text fits my reading purpose.	GLOB	0.53
7. I read slowly but carefully to be sure I understand what I'm reading.	PROB	0.54
8. I review the text first by noting its length and organizations.	GLOB	0.38
9. I try to get back on track when I lose concentration.	PROB	0.62
10. I underline or circle information in the text to help me remember it.	SUP	0.66
11. I adjust my reading speed according to what I'm reading.	PROB	0.64
12. When reading, I decide what to read closely and what to ignore.	GLOB	0.65
13. I use reference materials (e.g., a dictionary) to help me understand what I read	SUP	0.55
14. When text becomes difficult, I pay closer attention to what I'm reading.	PROB	0.64
15. I use tables, figures, and pictures in text to increase my understanding.	GLOB	0.64
16. I stop from time to time and think about what I'm reading.	PROB	0.67
17. I use context clues to help me better understand what I'm reading	GLOB	0.73
18. I paraphrase (restate ideas in Chinese) to better understand what I read.	SUP	0.63
19. I try to picture or visualize information to help remember what I read.	PROB	0.59
20. I use typographical features like boldface and italics to identify key information.	GLOB	0.61
21. I critically analyze and evaluate the information presented in the text.	GLOB	0.61
22. I go back and forth in the text to find relationships among ideas in it.	SUP	0.74
23. I check my understanding when I come across new information.	GLOB	0.76
24. I try to guess what the content of the text is about when I read.	GLOB	0.67
25. When text becomes difficult, I reread it to increase my understanding	PROB	0.68
26. I ask myself questions I like to have answered in the text.	SUP	0.52
27. I check to see if my guesses about the text are right or wrong.	GLOB	0.66
28. When I read, I guess the meaning of unknown words or phrases.	PROB	0.67
29. When reading, I translate from English into Chinese.	SUB	0.48
30. When reading, I think about information in both English and Chinese.	SUB	0.62
Variance accounted for: 37%		

* Hypothesized factors refer to the three subscales proposed by Mokhari & Reichard (2002a): global strategies, problem-solving strategies, and support strategies.

Table 2 presents the two-factor to four-factors solutions. The *two-factor solution* accounted for 40% of the variance; in other words, adding a second factor to the one-factor model only leads to a 3% increase in the explained variance. As for the factor structure, 28 items loaded on the first factor, and only item 29 and item 18 loaded on the second factor. Item 18 also cross-loaded on the first factor with a loading value of .34. It violated the simple structure criterion and did not meet the requirement of at least three indicators for a strong factor. Therefore, the two-factor solution was ruled out.

Table 2

Factor Analysis Results for SORS 2-Factor to 4-Factor Solutions (N = 715)

Item (abbreviated)	**Hypothesized Factors***	**Factor Loading (cutoff = .30)**								
		2-Factor		**3-Factor**			**4-Factor**			
		F1	**F2**	**F1**	**F2**	**F3**	**F1**	**F2**	**F3**	**F4**
1. a purpose in mind	GLOB	.55		.59						.34
2. take notes	SUP	.46		.55						.36
3. think about what I know	GLOB	.61		.75				.33		.42
4. take an overall view of the text	GLOB	.59		.30	.32			.35		
5. read aloud	SUP	.45								
6. check if text fits reading purpose	GLOB	.63		.52						.31
7. read slowly but carefully	PROB	.45		.53						
8. noting its length	GLOB	.32		.37				.38		
9. refocus	PROB	.50		.56				.49		
10. underline or circle information	SUP	.58		.57				.51		
11. adjust my reading speed	PROB	.62		.42				.79		
12. decide what to read and ignore	GLOB	.74		.36	.41			.63		
13. use reference materials	SUP	.43		.59						.36
14. pay closer attention	PROB	.53		.56					.30	.43
15. use tables, figures, and pictures	GLOB	.63		.57						.45
16. stop from time to time and think	PROB	.71		.49			.40			.40
17. use context clues	GLOB	.62		.46						
18. paraphrase (restate ideas in Chinese)	SUP	.34	.54	.38		.48			.62	
19. picture or visualize information	PROB	.74			.54		.63			
20. use typographical features	GLOB	.60		.36			31			
21. analyze & evaluate information	GLOB	.72			.57		.61			
22. go back and forth in the text	SUP	.74			.58		.51			
23. understanding about new information	GLOB	.65			47		.39		.33	
24. guess what the content of the text	GLOB	.55			.38				.33	
25. reread	PROB	.55		.32	31				.38	
26. ask myself questions	SUP	.57			.67		.55			
27. check guesses about the text	GLOB	.60			.81		.61			
28. guess the meaning of unknown words	PROB	.56			.41				.31	
29. translate from English into Chinese	SUB		.65			.64			.75	
30. think in English & Chinese	SUB	.53			.49		.34			
Variance Accounted for:		40%		43%			45%			

* Hypothesized factors refer to the three subscales proposed by Mokhari & Reichard (2002a): global strategies,

problem-solving strategies, and support strategies.

The *three-factor solution* failed to recover the theoretical structure of the three hypothesized subscales. It accounted for 43% of the variance, but it also did not achieve a simple structure. Most of the items loaded on the first two factors, and only item 29 and item 18 loaded on the third factor. Item 18 still presented a strong cross-loading on the first factor, and items 4, 12 and 25 all showed cross-loading issues. Besides, item 5 was not associated with any of the three factors. As a result, the three-factor solution was not considered a competitive one.

The *four-factor solution* accounted for 45% of the variance. Each factor was loaded by at least three items. In addition, the factor loadings of the items on the fourth factor were comparatively low (.31 ~ .45). Item 5 and item 7 were not associated with any of the four factors. Cross-loading problems were detected for item 3, item 14, and item 2. In addition, the interpretability of factors was ambiguous.

Therefore, the comparison of the four EFA solutions demonstrated that the general factor model already captured most of the variance that can be accounted for and adding extra factor (s) would not increase the explained variance that much or achieve a better factor structure. Hence only the general factor model was retained for the CFA analysis.

CFA Results

Based on previous studies and our EFA results, four alternative factor models were identified for initial CFA analysis: (1) general factor model (30 items), (2) reduced general factor model (14 items); (3) three-factor model which included "Global Reading Strategies" (13 items), "Problem Solving Strategies" (8 items), and "Support Strategies" (9 items); and (4) reduced three-factor model which included the following three factors "Global Reading Strategies" (5 items), "Problem Solving Strategies" (5 items), and "Support Strategies" (4 items).

CFA results were presented in Table 3. As shown in the table, all four models yielded a significant chi-square statistic, indicating an overall poor model fit to the sample data. It is worth noting that this absolute fit index is not accurate when a large sample is involved. Other fit indices suggested that the reduced three-factor model showed comparatively the best model fit (CFI = .904, TLI= .881, GFI = .925, SRMR = .049, RMSEA = .076) among the four models. The CFI, TLI, and GFI were not far away from the recommended value of .95; the SRMR was within the good fit range below the cutoff value .08. However, the high correlation among the latent factors (.873 ~ .931) indicated low discriminant validity among factors (Kline, 2016), which means the hypothesized three factors did not discriminate from each other, implying a higher-level common cause behind the three factors.

It was noted that there was only slight difference in the fit indices between general factor models and three-factor models when the number of indicators was accounted for. For example, the general model (CFI = 0.818, TLI= 0.805, GFI = 0.842, SRMR = 0.056, RMSEA = 0.079) showed similar model fit to the data as the three-factor model (CFI = 0.819, TLI= 0.804, GFI = 0.808, SRMR = 0.056, RMSEA = 0.079). However, a distinct difference in the model fit was detected between full models (30 items) and reduced models (14 items) after accounting for the number of factors. For example, the reduced general factor model showed great improvement in model fit (CFI = 0.904, TLI= 0.881, GFI = 0.925, SRMR = 0.049, RMSEA = 0.076) than the full general factor model (CFI = 0.818, TLI= 0.805, GFI = 0.842, SRMR = 0.056, RMSEA = 0.079). Suffice it to say, the model fit was more sensitive to the number of items of the SORS other than the number of factors, which supported a structure of less factors.

Table 3

Confirmatory Factor Analysis Fit Information for SORS Alternative Models

Fit Index	SORS CFA Models					Fit Criterion
	One-factor	One-factor (Reduced)	Three-Factor *Global/Problem/ Support*	Three-Factor (Reduced) *Global/Problem/ Support*	Revised One-Factor	
Chi-square	2230.41**	579.44**	2220.55**	328.72**	328.97**	x^2 (p > .05)
df	405	77	402	74	75	-
CFI	.818	.918	.819	.904	.960	> .95
TLI	.805	.903	.804	.881	.952	> .95
GFI	.842	.943	.808	.925	.944	> .95
SRMR	.056	.042	.056	.049	.031	< .08
RMSEA	.079	.068	.079	.076	.057	< .06
(90% CI)	.076 ~ .083	.062 ~ .073	.076 ~ .083	.069 ~ .084	.041 ~ .057	contains .05
Range of factor correlations			.282 ~ .299	.873 ~ .931		< .85

**p<0.01

Model Modification

Both the EFA and CFA results supported a unidimensional construct of the SORS, yet no model showed a satisfactory model fit. Therefore, we attempted to modify the full general factor model. Wu et al. (2012) and Mokhtari et al. (2018) both suggested the optimal number of indicators be about 15. The following steps were followed to drop items.

First, the ten items with the lowest loading values were removed. As all items are indicators of the general factor, removing the poor indicators improves the construct validity. Items 1, 2, 4, 5, 6, 7, 8, 13, 26, 29 were excluded for this process.

Second, the remaining 20 items were tested in CFA, and the residual matrix was examined. Residual correlations larger than 0.10 usually indicate there might be extra factors to extract. By this criterion, we found that the residuals were highly correlated between item 19 and item 16 (.107), item 20 (r = .125), item 21 (r = .195), item 25 (r = -.112); the residual of item 11 was highly correlated with item 10 (r = .16) and item 12 (r = .158); the residual of item 21 was highly correlated with item 18 (r = -.115) and item 20 (.099). Therefore, items 11, 19, and 21 were removed from the scale. The remaining 17 items were carefully interpreted in the context of L2 academic reading. Item 18 and item 30 refer to the use of participants' native language (Chinese) in reading; hence item 30 was dropped due to its lower loading.

As a final step, the one-factor model comprised of the final 16 items was again tested in CFA, and the modification indices were examined this time. The modification index (MI) value for a relationship between two variables indicates how much the model fit would improve if that relationship is included in the model. In terms of a general factor model, MI indicates the correlation between the residuals of the variables. All parameters with MIs > 10 were carefully reviewed. The largest MI value (*mi* = 54.54) was detected between item 14 and item 25. The contents of the two items were further examined. They all started with a similar wording "When text becomes difficult, …”. The correlation between the residuals of the two items could be the result of wording effects (Horan et al, 2003). Likewise, items 24, 27 and 28 all talked about guessing strategies; Items 17 and 18 shared the phrase “better understand what I read”. Therefore, besides the item-factor relation, four pairs of items were allowed to covary in the model: item14 with item 25, item 27 with item 28, item 17 with item 18, and item 24 with 28.

Eventually, our finalized scale retained 16 items (see Table 4). The general factor model of the revised SORS was tested in CFA and showed comparatively better model fit to the data (see Table 2, CFI = .960, TLI= .952, GFI = .944, SRMR = 0.031, RMSEA = 0.057), as all fit indices were within or close to the recommended range of good fit.

Table 4

The Revised SORS (16 items)

Item
3. I think about what I know to help me understand what I read.
9. I try to get back on track when I lose concentration.
10. I underline or circle information in the text to help me remember it.
12. When reading, I decide what to read closely and what to ignore.
15. I use tables, figures, and pictures in text to increase my understanding.
14. When text becomes difficult, I pay closer attention to what I’m reading.
16. I stop from time to time and think about what I’m reading.
17. I use context clues to help me better understand what I’m reading.
18. I paraphrase (restate ideas in Chinese) to better understand what I read.
20. I use typographical features like boldface and italics to identify key information.
22. I go back and forth in the text to find relationships among ideas in it.
23. I check my understanding when I come across new information.
24. I try to guess what the content of the text is about when I read.
25. When text becomes difficult, I reread it to increase my understanding.
27. I check to see if my guesses about the text are right or wrong.
28. When I read, I guess the meaning of unknown words or phrases.

Discussion

The study aimed to investigate the underlying latent structure of the survey of reading strategies (SORS). The SORS was assumed to have three subscales as the metacognitive awareness of reading strategies inventory (MARSI), for the SORS was adapted from the MARSI. The first question explored was whether the theoretical structure of the MARSI and SORS held across the sample.

In our study, the EFA analysis failed to recover the hypothesized three subscales. In the three-factor solution, most items were loaded on the first two factors, and only two items were loaded on the third factor. Some items also presented serious cross-loading issues. This is different from the findings of Mokhari and Reichard (2002a) and Wu et al (2012).

In the CFA analysis, neither the theoretical full (30-item) model (Mokhari & Reichard, 2002a) nor the reduced (14-item) model (Mokhtari et al., 2018) showed satisfactory fit to our sample data. The reduced three-factor model presented a comparatively better fit, yet the three factors in the model were highly correlated (.873 ~ .931), indicating a lack of discriminant validity between factors. In other words, the three factors most likely measure the same latent trait. This finding is consistent with Wu et al (2012)'s factor correlation results (.79 ~ .91).

The failure to recover the theoretical structure could be a result of sample bias or model misspecification. Our data were collected on Chinese college students via a snowball sampling method. A large number of observations were ruled out for validity concerns. Additionally, there were excessively more female respondents (1081) than male respondents (350). While the large sample size relieved some bias concerns, it was still possible that the factor analysis results were affected by the sample characteristics. However, sample bias does not rule out the possibility of model misspecification; that is, the hypothesized three-factor structure of the SORS or MARSI may not be the correct model, which brings the research question for further discussion about what the factor structure of the SORS would be.

While most researchers agreed upon the definition of metacognition as a two-dimension construct comprised of metacognitive awareness/knowledge and metacognitive regulation/monitor, this categorization was not well reflected in the development of metacognition scales. Prior studies have various, and sometimes conflicting opinions on the dimensions of metacognitive reading strategies. For example, *the Reading Strategy Use* (RSU, Pereira-Laird & Deane, 1997) consisted of metacognitive strategy use and cognitive strategy use. The *Metacognitive Reading Strategies Questionnaire* (MRSQ, Taraban, et al., 2004)) also measured two constructs: analytic cognitions aimed at reading comprehension and pragmatic behaviors aimed at studying and academic performance. The MARSI and SORS were made up of three subscales. The *Index of Reading Awareness* (IRA, Jacobs & Paris, 1987) and the *Metacognitive Reading Awareness Inventory* (MRAI, Miholic, 1994) both adopted four subscales: evaluation, planning, regulation and conditional knowledge.

Unlike all the opinions above, our comparison of one- through four-factor EFA solutions supported a general factor model. The CFA results also showed the three-factor models did not perform essentially better than the general factor models after accounting for the number of items. The highly correlated factors in the reduced three-factor model also suggested a common factor behind the hypothesized three factors. Therefore, the factor structure of the SORS should be at least unidimensional at a higher level. The disagreement on the subscales of metacognition of reading strategy among researchers suggests more factor analytic evidence are required to fully investigate the SORS construct. Future studies could focus on validation of the SORS with different samples of L2 English learners. Furthermore, special CFA models such as hierarchical models and bi-factor models (Kline, 2016) should be considered to account for the relationship between latent factors.

The last question refers to implications and suggestions for the use of the SORS in future studies. Our validation process did not provide construct validity evidence for its hypothesized theoretical structure. While score interpretation for individual items is still meaningful, researchers need to be cautious about interpreting the score of the three hypothesized subscales. Previous studies and our factor analytic results indicated a redundancy of items in the SORS. A revised SORS comprised of about 15 carefully selected items is sufficient for researchers to measure L2 learner's metacognitive reading strategy. However, revisions and modifications should be justified by theoretical support and practical meaning. Our revised SORS scale can serve as a reference.

Limitations

The latent structure of the SORS was investigated in this study, and some important findings were uncovered. However, the study was not flawless. First, the data for this study was collected via snowballing sampling. Although the participants were from different universities and colleges across China, they may not well represent the population of Chinese college students. Another concern refers to the missing observations. A total of 2,342 questionnaires were collected but only 1,431were considered valid after data cleaning. Although the sample size is sufficient for factor analyses to produce reliable results, the extent to which the sample characteristics impact the findings can hardly be evaluated. At the end of the study, the SORS was modified and showed a good fit to the data; however, scale revision criteria are not widely accepted. The revised SORS is only one of the possible alternative measures rather than the final solution.

Conclusion

The construct of metacognitive awareness and perceived use of reading strategies measured by the SORS is very likely unidimensional rather than multidimensional. Both the EFA and CFA results supported a general factor structure, which is different from the previous studies. As a result, the underlying structure of metacognition in L2 reading requires more factor analytic evidence and needs to be further explored with more samples of L2 learners.

The CFA results also suggested that the 30-item SORS overall performed no better than its shortened version. On the contrary, better model fit can be achieved by removing weak or overlapped items from the scale. However, the revision of the scale should be guided by strong theoretical considerations and practical meaning. A revised SORS with about 15 carefully selected items should be sufficient to assess L2 learners' use of metacognitive reading strategies.

References

Bandalos, D. L. (2018). *Measurement Theory and Applications for the Social Sciences*. New York & London. The Guilford Press.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1(2)*, 245–276. https://doi.org/10.1207/s15327906mbr0102_10.

Costello, A. B.; & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. Practical Assessment Research & Evaluation, 10(7), 1–9.

Cronbach, L. J. (1951). Coefficient Alpha and The Internal Structure of Tests. *Psychometrika, 16(3),* 297–334.

DiStefano, D. & Dombrowski, S. C. (2006). Investigating the Theoretical Structure of the Stanford-Binet–Fifth Edition*. Journal of Psychoeducational Assessment, 24(2),* 123–136. https://doi.org/10.1177/0734282905285244.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34(10),* 906–911. https://doi.org/10.1037/0003-066X.34.10.906.

Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.

Goodman, L. A. (1961). Snowball Sampling. *The Annals of Mathematical Statistics. 32(1),* 148–170.

Grabe, W. (2008). L1 and L2 reading relationships. In *Reading in a Second Language: Moving from Theory to Practice* (Cambridge Applied Linguistics, pp. 129–151). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139150484.011.

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.

Horan, P., DiStefano, C., & Motl, R. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling*, *10(3),* 435–455. https://doi.org/10.1207/S15328007SEM1003_6.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 32*, 179–185. https://doi.org/10.1007/BF02289447.

Miller, G. (2017). Metacognitive Awareness and Reading Strategy use: Investigating the Intermediate Level ESL Students' Awareness of Metacognitive Reading Strategies. *Culminating Projects in English. 115.* https://repository.stcloudstate.edu/engl_etds/115.

Hu, L & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modeling: A Multidisciplinary Journal*, *6 (1),* 1–55. https://doi.org/10.1080/10705519909540118.

Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist*, *22,* 255–278. https://doi.org/10.1080/00461520.1987.9653052.

Jou, Y. J. (2015). Investigation of Technological University Students' Use of Metacognitive Reading Strategies in First and Second Languages. English Language Teaching, 8(1), 180-188. http://dx.doi.org/10.5539/elt.v8n1p180.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141–151. https://doi.org/10.1177/001316446002000116.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Pres.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50 (9),* 741–749.

Miholic, V. (1994). An inventory to pique students' metacognitive awareness of reading strategies. *Journal of Reading*, *38 (2)*, 84–86.

Mokhtari, K., & Reichard, C. (2002a). Assessing Students' Metacognitive Awareness of Reading Strategies Inventory. *Journal of Educational Psycholog, 94 (2),* 249–259.

Mokhtari, K., & Sheorey, R. (2002b). Measuring ESL students' awareness of reading strategies. *Journal of Development Education*, *25 (3),* 2–10.

Mokhtari, K.; Dimitrov, D. M.; & Reichard, C. A. (2018). Revising the Metacognitive Awareness of Reading Strategies Inventory (MARSI) and testing for factorial invariance. *Education Faculty Publications and Presentations, 8(2),* 219–246.

Mulaik, S. A. (2010). *Foundations of factor analysis (2nd ed.)*. Boca Raton, FL: Taylor & Francis.

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

Pereira-Laird, J. A., & Deane, F. P. (1997). Development and validation of a self-report measure of reading strategy use. *Reading Psychology: An International Journal*, *18(3),* 185–235. https://doi.org/10.1080/0270271970180301.

Sheorey, R., and 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.

Smith, G. T. (2005). On Construct Validity: Issues of Method and Measurement. *Psychological Assessment. 17(4),* 396–408. https://doi.org/10.1037/1040-3590.17.4.396.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Taraban, R., Kerr, M., & Rynearson. (2004). Analytic And Pragmatic Factors In College Students' Metacognitive Reading Strategies. Reading Psychology, 25, 67–81. https://doi.org/10.1080/02702710490435547.

Tsai, Y. R.; Ernst, C.; & Talley, P. C. (2010). L1 and L2 Strategy Use in Reading Comprehension of Chinese EFL Readers. *Reading Psychology, 31(1),* 1–29. https://doi.org/10.1080/02702710802412081.

Van Kraayenoord, C. E. (2010). The role of metacognition in reading comprehension. In T. Hans-Peter., L. Wolfgang., & M. Peter (Eds.), *Focal Points of The Research and Development Of Pedagogically-Psychological Perspectives* (pp. 277–302). Hogrefe.

Victoria, Z. (2012). Awareness development for online reading. *Language Awareness*, *21(1–2),* 85-100. https://doi.org/10.1080/09658416.2011.639893.

Wu, L.; Valcke, M., & van Keer, H. (2012). Validation of a Chinese Version of Metacognitive Awareness of Reading strategies Inventory. *Egitim Arastirmalari – Eurasian Journal of Educational Research, 48,* 117–134.

Yaemtui, W. (2015). Investigating Reading Strategies Utilized by Able English Users and Less Able English Users of Thai EFL Students. *International Forum of Teaching and Studies. 11(1-2),* 55–68.

Zhang, L. (2018). *Metacognitive and Cognitive Strategy Use in Reading Comprehension.* Singapore. Springer Nature.

Zhang, L. J., & Wu, A. (2009). Chinese senior high school EFL students' metacognitive awareness and reading-strategy use. *Reading in a Foreign Language*, 21(1), 37–59.