

*Revising the Metacognitive Awareness
of Reading Strategies Inventory (MARSİ)
and testing for factorial invariance*

Kouider Mokhtari

The University of Texas at Tyler, USA

kmokhtari@uttyler.edu

Dimiter M. Dimitrov

National Center for Assessment in Riyadh, Saudi Arabia

George Mason University, USA

ddimitro@gmu.edu

Carla A. Reichard

The University of Texas at Tyler, USA

creichard@uttyler.edu

Abstract

In this study, we revised the *Metacognitive Awareness of Reading Strategies Inventory (MARSİ)*, a self-report instrument designed to assess students' awareness of reading strategies when reading school-related materials. We collected evidence of structural, generalizability, and external aspects of validity for the revised inventory (MARSİ-R). We first conducted a confirmatory factor analysis of the MARSİ instrument, which resulted in the reduction of the number of strategy statements from 30 to 15. We then tested MARSİ-R for factorial invariance across gender and ethnic groups and found that there is a uniformity in student interpretation of the reading strategy statements across these groups, thus allowing for their comparison on levels of metacognitive processing skills. We found evidence of the external validity aspect of MARSİ-R data through correlations of such data with a measure of the students' perceived reading ability. Given that this journal is oriented to second

language learning and teaching, our article also includes comments on the *Survey of Reading Strategies (SORS)*, which was based on the original MARSII and was designed to assess adolescents' and adults' metacognitive awareness and perceived use of ESL reading strategies. We provide a copy of the MARSII-R instrument and discuss the implications of the study's findings in light of new and emerging insights relative to assessing students' metacognitive awareness and perceived use of reading strategies.

Keywords: metacognition; reading comprehension; metacognitive awareness; reading strategies

1. Introduction

During the past two decades, reading researchers and practitioners have eagerly welcomed the re-emergence of scholarly interest in the role of metacognitive processing in students' reading comprehension performance. This renewed interest can be seen in the writing of several edited volumes devoted exclusively to the topic of metacognition (e.g., Garner, 1987; Hacker, Dunlosky, & Graesser, 1998; Hartman, 2001; Israel, Block, Bauserman, & Kinnucan-Welsch; 2005; Snow, 2002), the publication of a large number of articles addressing various aspects of metacognition and reading in scholarly journals, and the inclusion in several recently published books of instructional frameworks to guide the teaching of metacognitive reading strategies (e.g., Gersten, Fuchs, Williams, & Baker, 2001; Pearson & Gallagher, 1983; Pressley, 2000). Interest in the role of metacognition and reading is apparent in the publication of a special issue of the *International Electronic Journal of Elementary Education* (Desoete & Özso, 2009), and the launching of the *Metacognition and Learning* journal in 2006, with a special issue in 2011 (Schellings & van Hout Wolters, 2011) devoted exclusively to assessment and instructional issues pertaining to metacognition and reading. This article focuses on the following topics: (1) issues and primary purpose of the MARSII (Mokhtari & Reichard, 2003), (2) the validity of the MARSII, (3) a validity study using the MARSII-R, (4) discussion, and (5) comments on assessing metacognitive awareness and perceived reading strategy use of ESL students.

2. Issues and primary purpose of the MARSII

Despite the serious interest in metacognition and reading, an intricately connected web of issues and questions remains to be addressed prior to achieving a full understanding of the nature of the metacognitive processing skills and

strategies as they relate to reading and text understanding. This understanding should help in the design and development of adequate assessment measures of metacognitive reading strategies, as well as effective instructional and curriculum frameworks for advancing students' awareness and use of reading strategies when they read. Several contributors to the special issue of *Metacognition and Learning* published in 2011 (Schellings & van Hout Wolters, 2011) commented on the challenges and complexities related to metacognition and reading, in particular challenges related to the assessment of metacognitive processing strategies. In the following excerpt, MacNamara (2011) provides an excellent description of some of the potential challenges involved in "developing a pure (separable) measure of strategy use that is also reliable, valid, and contextualized" (p. 159):

There is a heightened understanding that metacognition and strategy use are crucial to deep, long-lasting comprehension and learning, but their assessment is challenging. First, students' judgments of what their abilities and habits are, and measurements of their performance often do not match. Second, students tend to learn and comprehend differently depending on the subject matter, contexts, goals, and tasks. As a consequence, a student may appear to use deep, reflective strategies in one situation, and fail to do so in other circumstances. Third, it is generally assumed that strategy use (metacognition, metacomprehension) are separable constructs from the underlying skills germane to the target task. (MacNamara, 2011, p. 159)

MacNamara's appraisal of the status of the field reminds us that, as a research community, we have a great deal more to do to develop adequate measures for assessing the cognitive and metacognitive processes involved in reading and text understanding.

We faced a number of theoretical, methodological, and practical challenges when we developed the original version of the MARSİ (Mokhtari & Reichard, 2002), which took a significant amount of time (nearly three years) and a great deal of effort on the part of several individuals. We were highly cognizant of the fact that it would be idealistic, and perhaps impractical, to try to develop a clean and discrete measure of strategy use that is also reliable, valid, and contextualized. As a result, we developed a reading strategy measure that was, by design, limited in terms of intended purpose, target audience, context or scope, and interpretation. At the request of teachers and researchers, we also developed the *Survey of Reading Strategies* (SORS; Mokhtari & Sheorey, 2002), an adapted version of the MARSİ instrument for use with learners of English as a second language (ESL). The MARSİ and SORS have been translated into several languages, specifically Arabic, Chinese, Czech, Farsi, French, German, Greek, Indonesian, Japanese, Korean, Polish, Slovenian, and Spanish. Both MARSİ and SORS have been widely used for teaching and research purposes, and published

in dozens of dissertations and other published research studies since their original publication in 2002.

The primary purpose of the MARSIs and SORS is to assess students' metacognitive awareness or perceived use of reading strategies when reading texts for academic purposes. When using self-report measures such as the MARSIs, it is important to consider the following two characteristics, which limit the interpretability of the results obtained from these measures. First, we designed the instrument to tap students' perceptions of reading strategy use (i.e., what strategies they think they use in general when reading), not actual strategy use (i.e., what specific strategies they actually used when reading). Second, we designed the instrument to tap students' strategy use in generalized reading contexts. In other words, when students complete the MARSIs, they are asked to think about a broad range of reading texts, tasks, or purposes, and, as a result, they report their perceived strategy use in a generalized rather than in a specified or contextualized sense.

We constructed the MARSIs so that we are able to uncover students' generalized use of reading strategies within the context of academic or school-related reading. In the instructions, we specifically ask students to select the strategies they believe they generally use when reading academic or school-related materials, as opposed to other types of reading materials (e.g., reading for pleasure). Given this context, it is also important to keep in mind that students' perceptions of strategy use are a reflection of a moment in time rather than a reflection of their reported strategies across different times, texts, or tasks.

Finally, we designed the MARSIs for a specific target audience, namely students with reading abilities that are roughly equivalent to those of a good reader in a typical upper elementary or middle grade classroom. Thus, we wanted to develop a measure that would enable us to identify student levels of metacognitive awareness or perceived use of reading strategies by reading ability rather than by grade level designation. There exists within any classroom or grade a range of readers and a range of reading ability levels. For example, in a fifth-grade classroom, there will be some readers who are as many as three grade levels ahead of the typical reader and some readers who are as many as three or more grade levels behind that benchmark.

3. Validity of the MARSIs

We used the unified construct-based model of validity (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Messick, 1989, 1995) to measure the validity of MARSIs. Under this model, there are six aspects of validity: (1) *content aspect of validity*, which includes evidence of content relevance, representativeness, and

technical quality; (2) *substantive aspect of validity*, which refers to theoretical rationales for the observed consistencies in item responses; (3) *structural aspect of validity*, which appraises the fidelity of the scoring structure to the structure of the construct domain at issue; (4) *generalizability aspect of validity*, which examines the extent to which score properties and interpretations generalize across population groups, settings, and tasks; (5) *external aspect of validity*, which includes convergent and discriminant evidence as well as evidence from measures of other traits; and (6) *consequential aspect of validity*, which relates to implications of score interpretations as a basis for action, as well as the actual consequences of test use, especially in regard to invalidity related to bias, fairness, and distributive justice (Messick, 1995; see also Dimitrov, 2012, pp. 41-51).

We obtained information about the validity of the original MARSİ instrument in stages. In the original study, we documented validity data, particularly on the content and substantive aspects of the instrument design and external evidence of correlation with reading ability. Subsequent studies have examined various aspects of the MARSİ, with many studies providing support for its validity, and a few raising issues pertaining to its appropriateness for college and adult readers and its association with reading ability (e.g., Guan, Roehrig, Mason, & Meng, 2010; MacNamara, 2007). Some issues are more difficult to address than others. As Cromley and Azevedo (2006), MacNamara (2011), as well as Veenman and colleagues (Veenman, 2011; Veenman, Van Hout-Wolters, & Afflerbach, 2006) have noted, self-report data have inherent limitations. There are methods of data collection (e.g., think-aloud protocols, reaction times, error detection, and other methods) that are less vulnerable to those limitations, but are also considerably more time-consuming and difficult to implement.

There are also issues with the generalized nature of the MARSİ directions: students use strategies to a different extent in different contexts, even in academic reading, and context-free measures do not accurately reflect strategy use for all of those contexts (e.g., Bråten & Strømsø, 2011; Hadwin, Winne, Stockley, Nesbit, & Woszczyna, 2001; Pressley, 2000; Veenman, 2011). However, contextualizing the instrument to focus on specific readings would necessarily limit its generalizability. An important aspect of validity that has not been thoroughly tested concerns the generalizability aspect of the MARSİ instrument. Characteristically, this question is addressed through testing for factorial invariance of the targeted construct across student populations, tasks, and contexts.

4. Validity study using the MARSİ-R

In light of the issues discussed above, we made a few changes to the MARSİ over the past several years, taking into account suggestions and recommendations

made by various researchers and practitioners who have used the instrument. These changes, which resulted in the MARSİ-R, pertain specifically to: (a) enhancements in the readability or comprehensibility of the strategy statements so that the instrument can be completed by students as early as fourth grade as long as they are able to read and understand the strategy statements; for example, a problem-solving strategy of “getting back on track when losing concentration” was revised as “getting back on track when sidetracked or distracted;” and (b) enhancements to the scale format and type of response expected to determine levels of strategy awareness or use, with the goal of improving the interpretation of the responses. The new 5-point scale taps students’ degree of knowledge and awareness of reading strategies ranging from “I have never heard of this strategy before” to “I know this strategy quite well, and I often use it when I read” (see Appendix for the MARSİ-R). While we do not expect these changes to significantly impact the overall factorial structure or reliability of the instrument, we believe this study is the first large-scale test of these changes.

The purpose of this study is to examine the factorial structure of the MARSİ in light of some changes in item wording and scale instructions and to collect evidence concerning the structural, generalizability, and external aspects of validity for the revised instrument (MARSİ-R). The tasks involved in addressing this goal relate to conducting confirmatory factor analysis of MARSİ-R data, testing for factorial invariance across gender and ethnic groups, and correlating MARSİ-R data with a relevant external criterion.

The testing of factorial invariance underlying students’ metacognitive awareness of reading strategies is of considerable practical importance for practitioners who wish to assess their students’ levels of metacognitive awareness of reading strategies and use the assessment data obtained to inform reading instruction. The generalizability of the instrument’s factor structures is also of considerable significance theoretically for researchers who are interested in studying differences in awareness or perceived use of reading strategies across different student populations and/or instructional interventions. Invariant or consistent factor structures would indicate that a level of uniformity in student interpretation of the reading strategy statements exists. In turn, this invariance makes it possible for us to compare student performance on metacognitive awareness measures, to develop a theoretical framework for guiding reading strategy instruction, and to determine the validity of assessment instruments when evaluating the quality of instruction.

4.1. Method

4.1.1. Participants

The participants in this study included 1,164 students in grades 6 through the first year of college. Students in grades 6-12 were enrolled in three large school districts and one community college located in a large metropolitan city in the Midwestern United States. The students ranged in age from 11 to 18 years old and the mean age of the group was 13.38 years ($SD = 1.99$). The sample included males (51%) and females (49%), representing a fairly diverse group with Caucasian ($N = 628$ or 54.0%), Hispanic ($N = 205$ or 17.6%), African-American ($N = 131$ or 11.2%), and Other ($N = 200$ or 17.2%) student groups. School demographics indicated that students were quite diverse with respect to linguistic, cultural, and socioeconomic backgrounds. For instance, Hispanic students had varied English language proficiency levels ranging from intermediate to advanced, as indicated by enrollment in either ESL and/or developmental reading classes. There were also discrepancies in socio-economic levels between minority student groups (i.e., Hispanics and African-American) and Caucasian students.

4.1.2. Instrument

All participants completed a modified version of a 30-item instrument – the *Metacognitive Awareness of Reading Strategies Inventory* (MARSi), which measures students' metacognitive awareness and use of reading strategies while reading academic materials. The modifications are described below.

The MARSi measures three broad categories of strategies including: (1) *global reading strategies* (GRS), which can be thought of as generalized, or global reading strategies aimed at setting the stage for the reading act (e.g., setting purpose for reading, previewing text content, predicting what the text is about, etc.); (2) *problem-solving strategies* (PSS), which are localized, focused problem-solving or repair strategies used when problems arise in understanding textual information (e.g., checking one's understanding upon encountering conflicting information, re-reading for better understanding, etc.); and (3) *support reading strategies* (SRS), which provide the support mechanisms or tools aimed at sustaining responsiveness to reading (e.g., the use of reference materials such as dictionaries and other support systems). These three classes of strategies interact with and support each other when used in the process of constructing meaning from text.

We validated the original MARSi instrument using large subject populations representing students with equivalent reading abilities ranging from middle school to college. Cronbach's coefficient alpha for internal consistency reliability of the three

documented subscales (global, problem-solving, and support reading strategies) ranged from .89 to .93, and score reliability for the total sample was .93, indicating reliable measures of metacognitive awareness of reading strategies. A complete description of the MARSIs, including its psychometric properties as well as its theoretical and research foundations, can be found in Mokhtari and Reichard (2002).

4.1.3. Data collection procedures

We collected the data during a three-week period of time during the spring semester of the school year. We administered the MARSIs-R in the English language to the subjects at the beginning of each class period, with the help of the classroom instructor, who was familiar with the tool and aware of the purpose of the study. After a brief overview of the objective of the study, a description of the instrument, and an explanation of the steps involved in completing it, the students were instructed to read each statement in the inventory and circle the number that best describes their responses to the statements. We advised students to work at their own pace, and reminded them to keep in mind reading academic or school-related materials while responding to the strategy statements. Finally, we let them know that there were no right or wrong responses to the statements, and that they could take as much time as they needed to complete the inventory. On average, the students completed the instrument in about 15-20 minutes.

4.1.4. Data analysis procedures

Given that there is prior theoretical and empirical work on establishing the underlying structure of the MARSIs (see Mokhtari & Reichard, 2002), we used a confirmatory factor analysis (CFA) to test the structural aspect of validity for MARSIs-R. We performed the CFA using the computer program for statistical analysis of latent variables Mplus (Muthén & Muthén, 1998-2012). Following the CFA, we tested the revised instrument (MARSIs-R) for factorial invariance across gender and ethnicity. After a preliminary analysis of the frequencies of responses across the five categories of the original ranking scale of MARSIs-R (see Appendix), the lowest two categories were collapsed, thus forming a 4-point ranking scale. This was done to stabilize the data in line with guidelines in the literature related to quality of rating scales (e.g., Dimitrov, 2012; Linacre, 2002).

The evaluation of data fit under the CFA in this study is based on a commonly used chi-square test statistic in combination with several other goodness-of-fit indices. An important clarification in this regard is that the CFA was conducted by using the computer program Mplus, which provides a dependable framework for analysis of categorically ordered data. The estimation of CFA parameters was

obtained through the use of a robust estimator for categorical data in Mplus, referred to as *weighted least square parameter estimates with standard errors and mean and variance adjusted* (WLSMV).

Evidence of data fit is provided when the chi-square value is *not* statistically significant ($p > .05$). However, given that the chi-square value rises with the increase of the sample size, which results in an artificial tendency to reject model fit, the evaluation of data fit is based on a joint examination of other goodness-of-fit indexes such as the *comparative fit index* (CFI), the *Tucker-Lewis index* (TLI), the *weighted root mean square residual* (WRMR), and the *root mean square error of approximation* (RMSEA) with its 90% confidence interval (CI). It should be clarified that the widely used *standardized root mean square residual* (SRMR) is appropriate for data on continuous variables and, therefore, not reported with Mplus analyses of categorical data; instead the WRMR index is provided.

Hu and Bentler (1999) suggested that a reasonably good fit is supported when the following fit criteria are met: CFI $\geq .95$, TLI $\geq .95$, and RMSEA $\leq .06$ (see also Bentler, 2004). Less stringent criteria of a reasonable data fit (CFI $\geq .90$, TLI $\geq .90$, and RMSEA $\leq .08$), can also be useful in some practical applications (e.g., Marsh, Hau, & Wen, 2004). The WRMR statistic is still viewed as an “experimental” fit index, with a value close to 1.0 indicating a good data fit at this stage of its use in CFA assessment of data fit (e.g., Cheung & Rensvold, 2002; Muthén & Muthén, 1998-2012.)

Under the original assignment of 30 items to three latent factors that were expected to underlie the responses on the MARSİ (global reading strategies, problem-solving strategies, and support reading strategies), we conducted CFA using Mplus with the WLSMV estimator for categorical variables.

5. Results

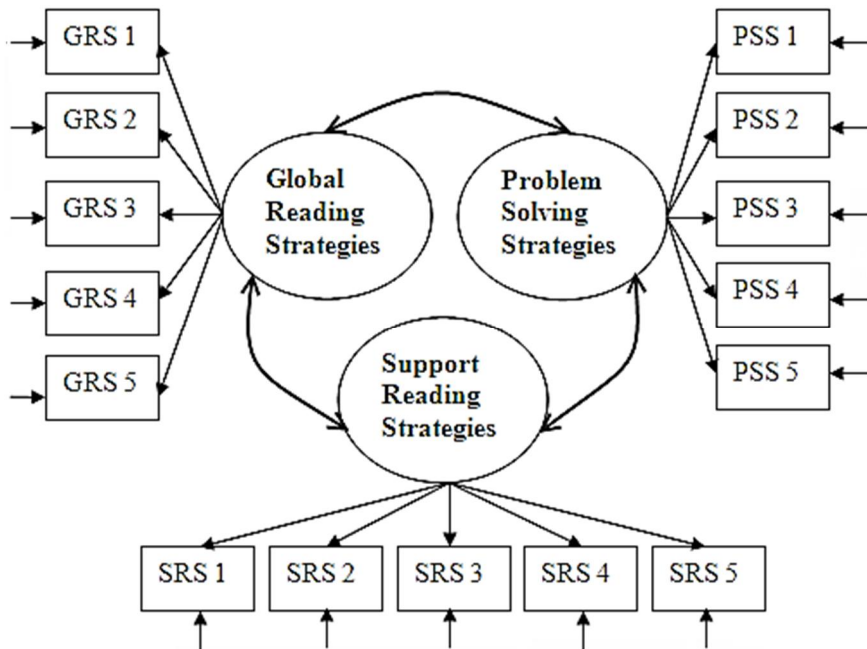
5.1. The model

The examination of the values for the goodness-of-fit indexes used in this study and the *modification indices* (MIs) reported in Mplus suggested the need for modification of the original factorial model for the MARSİ. For clarification, the MI value for a parameter gives the expected drop in the model chi-square value if this parameter is freely estimated (Jöreskog & Sörbom, 1979). Typically, MI greater than 10 (reported by default in Mplus) implies indication of misspecification for the respective parameter. In our case, although the estimates of the factor loading parameters for all items were statistically significant ($p < .001$), the MIs indicated numerous cross-loadings for items and correlated errors between items (not reported here for space considerations). Based on the examination of these misspecifications and related

substantive considerations, we modified the original MARSİ to the revised version, MARSİ-R, with five items per latent factor, for a total of 15 items.

Table 1 Description of the items associated with three latent factors under MARSİ-R

Factor/Item	Description
<i>Global reading strategies (GRS)</i>	
GRS 1:	Having a purpose in mind when reading
GRS 2:	Previewing text to see what it is about before reading
GRS 3:	Checking to see if the content of the text fits my purpose for reading
GRS 4:	Using typographical aids like bold face and italics to pick out key information
GRS 5:	Critically analyzing and evaluating the information read
<i>Problem-solving strategies (PSS)</i>	
PSS 1:	Getting back on track when getting sidetracked or distracted
PSS 2:	Adjusting my reading pace or speed based on what I'm reading
PSS 3:	Stopping from time to time to think about what I'm reading
PSS 4:	Re-reading to help ensure I understand what I'm reading
PSS 5:	Guessing the meaning of unknown words or phrases
<i>Support reading strategies (SRS)</i>	
SRS 1:	Taking notes while reading
SRS 2:	Reading aloud to help me understand what I'm reading
SRS 3:	Discussing what I read with others to check my understanding
SRS 4:	Underlining or circling important information in text
SRS 5:	Using reference materials such as dictionaries to support my reading



Note: GRS = global reading strategies, PSS = problem-solving strategies, SRS = support reading strategies

Figure 1 Baseline CFA model for MARSİ-R

The MARS-I-R model is described in Table 1 and graphically depicted in Figure 1. The means and standard deviations on the total MARS-I-R score by gender, ethnicity, and the total sample of 1,164 students are provided in Table 2. The goodness-of-fit indexes indicated an adequate data fit for this model. Specifically, although the chi-square value was statistically significant, $\chi^2(87) = 303.33$, $p < .001$, the other goodness-of-fit indexes suggested a good data fit, CFI = .972, TLI = .966, WRMR = 1.188, and RMSEA = .046, with 90%CI [.016 .027].

Table 2 Means and standard deviations of MARS-I-R scores by gender, ethnicity, and total sample

Group	<i>N</i>	<i>M</i>	<i>SD</i>
<i>Gender</i>			
Males	594	39.72	8.84
Females	570	42.36	8.46
<i>Ethnicity</i>			
African American	131	41.96	8.13
Hispanic Latino	205	39.99	8.90
Caucasian	628	41.35	8.74
Other	200	40.41	8.94
Total	1,164	41.01	8.75

Note. *N* = Sample size, *M* = Mean, *SD* = Standard deviation

Table 3 Standardized estimates of factor loadings for the baseline CFA model

Factor/items	Factor loading estimate (FLE)	Standard error (SE)	Significance <i>p</i> -value
<i>Global reading strategies</i>			
GRS 1	0.549	0.027	< .001
GRS 2	0.584	0.026	< .001
GRS 3	0.647	0.025	< .001
GRS 4	0.632	0.026	< .001
GRS 5	0.670	0.024	< .001
<i>Problem-solving strategies</i>			
PSS 1	0.606	0.025	< .001
PSS 2	0.666	0.022	< .001
PSS 3	0.667	0.021	< .001
PSS 4	0.594	0.025	< .001
PSS 5	0.520	0.026	0.000
<i>Support reading strategies</i>			
SRS 1	0.560	0.025	< .001
SRS 2	0.532	0.025	< .001
SRS 3	0.678	0.021	< .001
SRS 4	0.692	0.020	< .001
SRS 5	0.725	0.021	< .001

Furthermore, as shown in Table 3, the estimates of the standardized factor loadings for all items are sizable (> .40) and statistically significant ($p < .001$).

The correlations among the factors under MARSIR were found to be: (1) $r = .814$ between *global reading* and *problem-solving strategies*, (2) $r = .618$ between *global reading* and *support reading strategies*, and (3) $r = .840$ between *problem-solving strategies* and *support reading strategies*.

5.2. Reliability

Cronbach's alpha coefficient for internal consistency reliability of the 15-item scale MARSIR was equal to .850. By subscales, the alpha values for *global reading strategies*, *problem-solving strategies*, and *support reading strategies* were .703, .693, and .743, respectively. These relatively low estimates of internal consistency reliability of the three subscales are partly due to the smaller number of subscale items (five items per subscale).

5.3. Convergence with external measures

As a part of collecting evidence related to the external aspect of validity, we correlated the subscale scores and the total scores on the MARSIR with the scores on the variable *reader*. The variable *reader*, which asks students to estimate their level of reading ability, comes from the *General information* section of the MARSIR and represents respondents' answer to the item: "I consider myself: (1) an excellent reader, (2) a good reader, (3) an average reader, or (4) a poor reader." We found the correlation coefficients, all statistically significant ($p < .001$), to be (1) $r = .316$ between *reader* and *global reading strategies*, (2) $r = .346$ between *reader* and *problem-solving strategies*, (3) $r = .163$ between *reader* and *support reading strategies*, and (4) $r = .330$ between *reader* and the total scale score on the MARSIR. Regarding the relationship between the student grade level and scale scores on the MARSIR, the only statistically significant, yet low, correlation was between the grade level of the students and their score on the subscale *global reading strategies* ($r = .08$, $p = .009$). An overall implication of this finding is that the grade level of the students is unrelated to their relative performance on the MARSIR.

5.4. Results of testing for factorial invariance of the MARSIR across gender and ethnicity

Testing for factorial invariance of a CFA model across gender and ethnicity is conducted to examine the extent to which the three-factor structure of the MARSIR and the score interpretations generalize across gender and ethnic groups. That is, the question is whether the MARSIR's underlying construct has

the same meaning across the gender and ethnic groups in this study. To clarify some basic terms, *configural invariance* refers to invariance of the model configuration across the respective groups (e.g., males and females). *Measurement invariance* refers to: (1) metric invariance – equal factor loadings across groups, (2) scalar invariance – equal item intercepts across groups, and (3) invariance of item uniquenesses – equal item residual variances/covariances across groups. *Structural invariance* refers to invariance of factor variances and covariance (e.g., Byrne, 1988; Byrne, Shavelson, & Muthén, 1989; Dimitrov, 2012).

We performed the testing for factorial invariance using the *step-up constraints method*. Under this approach, the analysis begins with the least constrained solution (total lack of invariance) and subsequent restrictions for equality of specific parameters across groups are imposed, thus producing nested models that are tested against each other using the chi-square difference test. It should be emphasized, however, that by using the WLSMV estimator in CFA with categorical variables, the conventional approach of taking the difference between the chi-square values and the difference in the degrees of freedom is *not* appropriate because the chi-square difference is not distributed as chi-square. Therefore, the DIFFTEST option in Mplus was used here to conduct chi-square difference tests in the comparison of nested CFA models under WLSMV estimation with categorical variables (Muthén & Muthén, 1998-2012, p. 625).

Table 4 Configural invariance of the CFA baseline model of across gender and ethnicity

Group	χ^2	df	p	90% CI for RMSEA					
				CFI	TLI	WRMR	RMSEA	LL	UL
Gender									
Male	184.88	87	.000	0.974	0.969	0.932	0.044	0.035	0.052
Female	213.05	87	.000	0.966	0.959	1.014	0.050	0.042	0.059
Ethnicity									
Caucasian	258.26	87	.000	0.963	0.955	1.095	0.056	0.048	0.064
Other	175.93	87	.000	0.972	0.966	0.923	0.044	0.034	0.053
Total	303.33	87	.000	0.972	0.966	1.188	0.046	0.041	0.052

Note. CI = confidence interval; LL = lower limit; LU = upper limit

5.4.1. Factorial invariance across gender

To test for configural invariance across gender, we first tested the MARS-I-R model in Figure 1 for data fit separately for males and females. The results in Table 4 indicate there is a good data fit across males and females, as well as for the total sample of respondents, thus supporting the configural invariance of the MARS-I-R model referred to hereafter also as a *baseline model*. The correlations among the latent factors *global reading strategies*, *problem-solving strategies*,

and *support reading strategies*, obtained with the baseline model for the total sample ($N = 1,164$) were quite strong, namely (1) .814 between GRS and PSS, (2) .618 between GRS and SRS, and (3) .840 between PSS and SRS.

Table 5 Testing for factorial invariance of the MARS-R model across gender

Model	χ^2	df	Comparison	DIFF χ^2	Δ df	p -value
Model 0	398.210	174		—	—	—
Model 1	389.252	186	M1-M0	22.288	12	.034
Model 1P	373.423	185	M1P-M0	13.150	11	.284
Model 2	446.064	212	M2-M1P	53.209	27	.002
Model 2P	432.259	210	M2P-M1P	34.891	25	.090
Model 3	441.655	225	M3-M2P	26.791	15	.030
Model 3P	437.756	224	M3P-M2P	21.902	14	.081
Model 4	415.628	227	M4-M3P	1.234	3	.745
Model 5	379.257	230	M5-M4	2.058	3	.560

Note. M0 = Model 0: Baseline model (no invariance); M1 = Model 1: Invariance of factor loadings; M1P = Model 1P: Partially invariant factor loadings; M2 = Model 2: Model 1P + invariant thresholds; M2P = Model 2P: Model 1P + partially invariant thresholds; M3 = Model 3: Model 2P + invariant item residual variances; M3P = Model 3P: Model 2P + partially invariant item residual variances; M4 = Model 4: Model M3 + invariant factor variances; M5 = Model 5: Model 4 + invariant factor covariances; with model comparisons, the respective hypothesis of invariant parameters across gender is retained when the DIFF χ^2 value is *not* statistically significant ($p > .05$)

The results from testing for measurement and structural invariance of the baseline model are summarized in Table 5, where subsequent pairs of nested models are tested against each other using the DIFFTEST option in Mplus for chi-square difference tests with categorical variables.

Model 1 is obtained from the baseline model (Model 0) by imposing the constraint of invariant factor loadings (Model 1 is nested within Model 0). As the DIFF χ^2 is statistically significant ($p = .034$), not all factor loadings are invariant across males and females. The examination of the modification indices (MIs) showed that the factor loading of one item (PSS5) associated with the factor *problem-solving strategies* is not invariant across gender. After relaxing the constraint for invariant loading for this item, which resulted in a model denoted Model 1P, the comparison of Model 0 versus Model 1P produced a nonsignificant DIFF χ^2 value ($p = .284$). This indicated the presence of a partial invariance for factor loadings across gender – except for item PSS5, the factor loadings are invariant across males and females.

Next, Model 2 is obtained from Model 1P by imposing invariance of the item thresholds (latent cutting values between adjacent response categories on MARS-R items) across gender. Thus, Model 2 is nested within Model 1P. As the DIFF χ^2 for the comparison of Model 2 versus Model 1P is statistically significant ($p = .002$), not all item thresholds are invariant across males and females. After

examining the modification indices (MIs) and successive free estimation of thresholds, a nonsignificant $\text{DIFF}\chi^2$ value was obtained with Model 2P in which two thresholds were freely estimated (i.e., noninvariant across gender), that is, the thresholds between the first two response categories for items GRS1 and GRS5 (see Table 1 for description). Thus, we have established that there is a partial invariance of item thresholds across gender, with two (out of 45 thresholds in total) being different across males and females.

As a next step, Model 3 was developed from Model 2P by imposing invariance of the item residual variances across gender. The $\text{DIFF}\chi^2$ for the comparison of Model 3 as nested within Model 2P was statistically significant ($p = .030$), thus indicating that there is no full invariance of item residual variances across gender. After examining the modification indices (MIs) and freely estimating the residual variance for one item (GRS2), the $\text{DIFF}\chi^2$ for the comparison of the resulting Model 3P versus Model 2P was no longer statistically significant ($p = .081$). Thus, there is a partial invariance of item residual variances, with the residual variance of one item (out of 15) being noninvariant across gender.

Model 4 was obtained from Model 3P by imposing of invariance of the factor variances across gender. The $\text{DIFF}\chi^2$ for the comparison of Model 4 as nested within Model 3P was not statistically significant ($p = .745$), thus indicating the variances of the three latent factors were the same for males and females. Finally, Model 5 was obtained from Model 4 by imposing invariance of the covariances among the latent factors. The $\text{DIFF}\chi^2$ test comparing Model 5 as nested within Model 4 was not statistically significant ($p = .560$), thus indicating the covariances among the three latent factors do not change across males and females.

5.4.2. Factorial invariance across ethnic groups

As 54% of the total sample were Caucasian students (see Table 2), the testing for factorial invariance across ethnic groups was conducted by comparing Caucasian versus non-Caucasian groups of students. The results are summarized in Table 6. Following the procedure of sequential comparisons of nested models, described in detail with the testing for factorial invariance across gender, it was found that: (1) all factor loadings were invariant, (2) the item thresholds were invariant, with the exception of the second threshold of two items (GRS4, GRS5) and the third threshold of two items (GRS4 and SRS2), (3) the item residual variances were invariant, with the exception of five items (GRS5, PSS1, PSS2, SRS1, and SRS4), (4) the variances of the latent factors, *problem solving strategies* and *support reading strategies* were invariant, but not the variance of the *global reading strategies*, and (5) the covariances among the three latent factors were invariant across the ethnic groups.

To summarize the results in this section, noninvariance across gender was signaled for the factor loading of one item (PSS5), the thresholds between the first two response categories for two items (GRS1 and GRS5), and the residual variance for one item (GRS2). Given that up to 20% noninvariant parameters are tolerable for an acceptable partial invariance (e.g., Dimitrov, 2012), the conclusion is that there is a satisfactory level of partial measurement invariance across gender for the MARSİ-R. At the same time, the variances of all three latent factors and the covariances among them were found invariant, thus indicating full structural invariance of MARSİ-R across gender. Regarding the two ethnic groups used in this study (Caucasian and Non-Caucasian), it was found that all factor loadings were invariant, whereas noninvariance was signaled for: (1) the second threshold of two items (GRS4, GRS5) and the third threshold of two items (GRS4 and SRS2), and (2) the variance of one latent factor (GRS). The covariances among all three latent factors were invariant across the two ethnic groups. The conclusion is that there is an acceptable level of partial measurement and structural invariance of the MARSİ-R across the two ethnic groups.

5.5. Testing for gender and ethnic differences on MARSİ-R factors

Given the presence of an adequate factorial invariance across gender and ethnicity for the MARSİ-R data, testing for gender and ethnic differences on the latent factors of MARSİ-R (*global reading strategies*, *problem-solving strategies*, and *support reading strategies*) is appropriate. Such testing was conducted by regressing each of these three factors on gender and ethnicity in the baseline CFA model for MARSİ-R (see Figure 1). For gender, the regression coefficients on the three latent factors are denoted here as γ_1 , γ_2 , and γ_3 , respectively, whereas the regression coefficients for ethnicity on the latent factors are denoted β_1 , β_2 , and β_3 , respectively.

Regarding gender, the estimates of all regression coefficients were statistically significant, with their magnitudes, p -values, and effect size, d , being (a) $\gamma_1 = .120$, $p = .008$, $d = .185$, (b) $\gamma_2 = .240$, $p < .001$, $d = .320$, and (c) $\gamma_3 = .205$, $p < .001$, $d = .312$. The effect size estimate, d , indicates how many latent standard deviations separate the means of males and females on the factor of interest (Hancock, 2004). Under Cohen's (1988) interpretation for the magnitude of effect size, there is a small effect size for the gender difference in favor of females on each of the three latent factors, with the relatively largest effect size being on *problem-solving strategies*, followed by the effect size for *support reading strategies* and *global reading strategies*.

Table 6 Configural invariance of the CFA baseline model of the MARSİ-R across ethnic groups (Caucasian, non-Caucasian)

Model	χ^2	df	Comparison	DIFF χ^2	Δdf	p-value
Model 0	431.174	174		—	—	—
Model 1	403.598	186	M1-M0	16.931	12	.152
Model 2	499.183	213	M2-M1	74.155	27	.000
Model 2P	458.214	209	M2P-M1	27.436	23	.239
Model 3	486.372	224	M3-M2P	42.178	15	.000
Model 3P	459.681	219	M3P-M2P	16.380	10	.089
Model 4	459.781	222	M4-M3P	8.241	3	.041
Model 4P	458.281	221	M4P-M3P	5.401	2	.067
Model 5	416.884	224	M5-M4P	1.462	3	.691

Note. M0 = Model 0: Baseline model (no invariance); M1 = Model 1: Invariant factor loadings; M1P = Model 1P: Partially invariant factor loadings; M2 = Model 2: Model 1P + invariant thresholds; M2P = Model 2P: Model 1P + partially invariant thresholds; M3 = Model 3: Model 2P + invariant item residual variances; M3P = Model 3P: Model 2P + partially invariant item residual variances; M4 = Model 4: Model M3 + invariant factor variances; M4P = Model 4P: Model 3 + partially invariant factor variances; M5 = Model 5: Model 4P + invariant factor covariances; with model comparisons, the respective hypothesis of invariant parameters across gender is retained when the DIFF χ^2 value is *not* statistically significant ($p > .05$)

Regarding ethnicity, the estimate of the regression coefficient for *global reading strategies* was statistically significant ($\beta_1 = -.125$, $p = .006$, $d = .192$), thus indicating a small effect size of the ethnic difference on *global reading strategies* in favor of the Caucasian students (the data coding for ethnicity is 0 = Caucasian, 1 = non-Caucasian). There was no statistical significance for the estimates of regression coefficients on the other two latent factors, thus indicating a lack of ethnic differences on *problem-solving strategies* ($\beta_2 = -0.022$, $p = .683$) and *support reading strategies* ($\beta_3 = 0.038$, $p = .397$).

5.6. Correlations between MARSİ-R latent factors and students' perceived reading ability

In search of evidence related to the external aspect of validity, we incorporated the variable *reader* in the MARSİ-R model depicted in Figure 1 to examine its correlational relationships to the three latent factors. The estimates of correlations between the students' scores on *reader* and their latent (true-score) performance on strategies of *global reading*, *problem-solving*, and *support reading*, all statistically significant ($p < .0001$), are reported in Table 7. As can be seen, these correlation estimates are higher than their counterparts, reported earlier in this paper, when the raw scores on the three factors are used: (1) .373 versus .316, (2) .419 versus .346, and (3) .190 versus .163 for *global reading*, *problem-solving*, and *support reading*, respectively. This is due to attenuation of the correlations when raw scores (instead of true scores) are used.

Table 7 Correlations among latent scores on the MARSİ-R subscales, grade level of the students, and their perceived level of reading ability

Variable	GRS	PSS	SRS	Grade	Reader
GRS		.815***	.619***	.082*	.373***
PSS			.840***	.061	.419***
SRS				.048	.190***
Grade					.042

Note. GRS = global reading strategies, PSS = problem-solving strategies, SRS = support reading strategies, Grade = grade level, reader = students' self-perception of their reading ability (an external measure in MARSİ-R); * $p < .05$. ** $p < .01$. *** $p < .001$

Table 7 also provides correlations between the grade level of the students and their latent (true-score) performance on strategies of *global reading*, *problem-solving*, and *support reading*. These estimates were obtained by incorporating correlations between the grade level variable and the three latent factors in the CFA model depicted in Figure 1. Statistically significant, yet very small, is only the correlation between grade level and *global reading* ($r = .082$, $p < .05$). These results are consistent with the correlations between grade level and the raw scores on *global reading*, *problem-solving*, and *support reading* reported earlier in this paper. An overall implication of this finding is that the grade level of the students is unrelated to their relative performance on the MARSİ-R.

6. Discussion

In this study, we revised the original MARSİ and collected evidence of structural, generalizability, and external aspects of validity for the revised inventory (MARSİ-R). We first conducted a confirmatory factor analysis of the revised MARSİ, which resulted in the reduction of the number of strategy statements from 30 to 15. This result occurred because some strategy statements appeared to tap similar reading strategy constructs (see Appendix). We subsequently tested the MARSİ-R for factorial invariance across gender and ethnic groups and found that there is a uniformity in student interpretation of the reading strategy statements across these groups, thus allowing for their comparison on levels of metacognitive processing skills. Finally, we found evidence of the external validity aspect of MARSİ-R data through correlations of such data with a measure of the students' perceived reading ability.

The results from the confirmatory analysis of MARSİ-R data supported the original factorial structure of three latent factors – *global reading strategies* (GRS), *problem-solving strategies* (PRS), and *support reading strategies* (SRS) – with five reading strategy statements serving as indicators for each latent factor.

The internal consistency reliability of the student scores on the indicators by latent factors was reasonably high.

Furthermore, we found the factorial structure of the MARSIR-R to be invariant across gender and ethnic groups, namely Caucasian versus other ethnic groups taken together for sample consideration (Hispanic, African-American, and Other). This finding indicates that, regardless of gender and ethnicity, the students assign the same meanings to the reading strategy statements in the inventory. Therefore, it is appropriate to compare gender and ethnic groups on their performance on the MARSIR-R. It also found that the relative performance of the students on the MARSIR-R does not depend on their grade level. These findings are helpful when exploring differences in metacognitive awareness or perceived use of reading strategies across student populations, for developing instructional frameworks and curriculum materials aimed at enhancing students' levels of metacognitive processing strategies, and for determining the validity of metacognitive assessment instruments when evaluating the quality of instruction.

In relation to the validity of the MARSIR-R, the results in this study provide evidence about: (1) the structural aspect of validity, with a three-factor structure (GRS, PSS, and SRS), (2) the generalizability aspect of validity, with factorial invariance across gender and ethnic groups, and (3) the external aspect of validity, with correlations between the students' scores on each of the three MARSIR-R subscales (GRS, PSS, SRS) and their scores on the *reader* scale as an external measure of perceived reading ability.

The factorial invariance of MARSIR-R data across gender and ethnicity has both theoretical and practical implications. Theoretically, the generalizability of the MARSIR-R's latent factor structure indicates that there is uniformity in student interpretation of the reading strategy statements. This makes it possible for researchers to design studies aimed at exploring student awareness of reading strategies across student populations, to develop theoretical frameworks for understanding student metacognitive awareness of reading strategies in relation to reading comprehension performance, to design instructional interventions aimed at enhancing student metacognitive awareness and use of reading strategies when reading, and to determine the validity of measures such as the MARSIR-R when evaluating the quality of instruction.

The generalizability of the MARSIR-R's factor structure is also of considerable practical importance to classroom teachers, reading specialists, and other education professionals who are interested in identifying measures for reliable and valid assessment of students' metacognitive awareness of reading strategies. A useful practical implication of the consistent factor structure of the MARSIR-R is that student ratings of their perceived awareness or use of reading

strategies are not affected by bias arising from differences in interpretation of the same scales in the inventory across different student populations.

However, it is important to note that the generalizability of the factor structure of the MARSİ-R has limits that need to be addressed through further research and exploration of students' judgments of their perceived metacognitive awareness or use of reading strategies when reading. While we found that the understanding of the students about the levels of their metacognitive awareness is consistent across gender and ethnic groups within a set of school districts in one metropolitan area, we are not certain that a similar level of invariance would be found across more disparate groups of students in more or less linguistically and culturally diverse school settings.

A note on correlations of the MARSİ-R with reading ability is also warranted. Specifically, one of the persistent issues with the MARSİ has been the relatively low correlations between reported scores of strategy use on the MARSİ with external measures of reading ability. If strategy usage is important to reading comprehension, we would expect to see higher correlations. Undoubtedly, the issues discussed here with regard to self-report instruments, generalized (vs. contextualized) usage, and so forth, play a role in this correlation problem.

We want to mention yet another, previously unpublished, issue which came up in the testing of the original MARSİ instrument. Specifically, in the initial pilot testing of the MARSİ, we began with 60 items, which were then winnowed down to 30. In the initial analysis of 60 items, there was one item ("When reading difficult materials, I give up") which, when grouped with some of the items later included in the *support reading strategies* factor, had a significant negative correlation with self-reported reading ability. We omitted this item from the published original version of the MARSİ because it did not lend itself to any specific instructional strategy or specific theoretical finding. However, because of its strong (negative) correlation with reading ability, we note it here for any researchers who may be interested in pursuing it. We recognize that in many cases, struggling readers may be unable to adequately diagnose their own deficiencies in detail, though they do know that they feel like just giving up.

Analysis of the results of this study leaves us in a good position to consider important questions and issues that might be addressed in future studies. First, we want to reiterate our cautions related to the uses and interpretation of the results obtained from this shorter, revised version of the MARSİ. We ask that MARSİ-R users keep in mind the fact that this instrument asks students to rate their strategy use in a generalized rather than a specific, contextualized sense. Second, we encourage researchers and practitioners to use the MARSİ-R in their work to determine the extent to which it provides useful information for deter-

mining students' levels of metacognitive processing. Third, we hope that researchers would consider carrying out cross-text, cross-task, and cross-language comparisons of instruction in metacognitive awareness, as such studies would help us to better understand whether and to what extent students' metacognitive awareness and use of reading strategies are text-specific, task-specific, or language-specific. Findings of such studies may also help us to determine why there are so few significant effects of metacognitive awareness on measures of reading comprehension.

7. Final comments on assessing metacognitive awareness and perceived reading strategy use of ESL students

Given that this journal is oriented toward second language learning and teaching, it is important to comment here on assessing the metacognitive awareness and perceived reading strategy use of ESL students as well. As indicated in the first section of this manuscript, we developed an adapted version of the original MARSİ instrument for use with ESL students and we called it the *Survey of Reading Strategies* (Mokhtari & Sheorey, 2002) to distinguish it from the MARSİ, although the MARSİ and SORS are similar in terms of design and implementation features.

Like the MARSİ, the SORS is a self-report instrument aimed at assessing students' metacognitive awareness and use of reading strategies when reading academic or school-related materials. In this adapted version, we made slight revisions to a few of the strategy statements with the goal of improving their comprehensibility for ESL students. For instance, we revised the instructions for administration as well as interpretation of the results for clarity and readability purposes. In addition, we integrated certain ESL reading strategies (e.g., use of cognates, code-mixing or code-switching, and translation across two or more languages) that are characteristically used by bi-literate or multi-literate readers when reading academic texts in English.

It is worth noting that both the MARSİ and SORS are valid measures for assessing students' metacognitive awareness and perceived use of reading strategies. Information about the development of the MARSİ and SORS instruments, their psychometric properties, as well as their limitations can be found in Mokhtari and Reichard (2002), and Mokhtari and Sheorey (2002). Although the SORS is a valid instrument, we intend to revise and revalidate it to follow the practice of the MARSİ as well. The decision as to which measure to use depends to a large extent on the students' levels of English proficiency. For students with advanced levels of English proficiency, either measure is fine to use. However, there is practical value in using the SORS when assessing students with lower levels of English proficiency.

The MARSİ and the SORS have been translated into several languages with translations used for students representing different linguistic and cultural backgrounds. Translated versions of MARSİ and SORS are available in Arabic, Chinese, Czech, Farsi, French, German, Greek, Indonesian, Japanese, Korean, Polish, Slovenian, and Spanish. Both MARSİ and SORS have been widely used around the globe by classroom teachers and researchers with students varying in levels of language proficiency. A number of studies using either the MARSİ or the SORS have been published as master's or doctoral dissertations and in refereed journals.

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bentler, P. M. (2004). *EQS 6: Structural equation program manual*. Encino, CA: Multivariate Software.
- Bråten, I., & Strømsø, H. L. (2011). Measuring strategic processing when students read multiple texts. *Metacognition and Learning, 6*, 111-130.
- Byrne, B. M. (1988). The self-description questionnaire III: Testing for equivalent factorial validity across ability. *Educational and Psychological Measurement, 48*, 397-406.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cromley, J. G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning, 1*, 229-247.
- Desoete, A., & Özso, G. (Eds.). (2009). Metacognition [Special issue]. *International Electronic Journal of Elementary Education, 2*(1).
- Dimitrov, D. M. (2012). *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria, VA: American Counseling Association.
- Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.
- Gersten, R., Fuchs, L., Williams, J., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research, 71*, 279-320.
- Guan, C. Q., Roehrig, A. D., Mason, R., & Meng, W. (2010). Psychometric properties of *Meta-cognitive Awareness of Reading Strategy Inventory*. *Journal of Educational and Developmental Psychology, 1*, 3-17.
- Hacker, D. J., Dunlosky, J., & Graesser, A. C. (1998). *Metacognition in educational theory and practice*. Mahwah, NJ: Lawrence Erlbaum.
- Hadwin, A.F., Winne, P.H., Stockley, D. B., Nesbit, J., & Woszczyzna, C. (2001). Context moderates students' self-reports about how they study. *Journal of Educational Psychology, 93*, 477-487.

- Hancock, G. R. (2004). Experimental, quasi-experimental, and non-experimental design and analysis with latent variables. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 317-334). Thousand Oaks, CA: Sage.
- Hartman, H. J. (Editor). (2001). *Metacognition in learning and instruction: Theory, research and practice*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Hu, L.T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.
- Israel, S. E., Block, C. C., Bauserman, K. L., & Kinnucan-Welsch, K. (Eds.). (2005). *Metacognition in literacy learning: Theory, assessment, instruction, and professional development*. Mahwah, NJ: Lawrence Erlbaum.
- Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.
- MacNamara, D. S. (2007). *Reading comprehension strategies: Theories, interventions, and technologies*. Mahwah, NJ: Lawrence Erlbaum.
- MacNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: challenges and successes. *Metacognition and Learning, 6*, 195-203.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cut-off values for fit indexes and dangers in overgeneralizing findings. *Structural Equation Modeling, 11*, 320-341.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: American Council on Education.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.
- Mokhtari, K., & Reichard, C. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*(2), 249-259.
- Mokhtari, K., & Sheorey, R. (2002). Measuring ESL students' awareness of reading strategies. *Journal of Developmental Education, 25*(3), 2-10.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Pearson, P. D., & Gallagher, M. C. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology, 8*, 317-344.
- Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research*, (Vol. 3, pp. 545-561). Mahwah, NJ: Lawrence Erlbaum.

- Schellings, G., & van Hout Wolters, B. (2011). Measuring strategy use with self-report instruments [Special issue]. *Metacognition and Learning*, 6(1).
- Snow, C. (2002). *Reading for understanding: Towards an R&D program in reading comprehension*. Santa Monica, CA: Rand Education.
- Veenman, M. V. J. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning*, 6, 205-211.
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1, 3-14.

APPENDIX

METACOGNITIVE AWARENESS OF READING STRATEGIES INVENTORY-REVISED (MARSIR, 2013)

GENERAL INFORMATION

Name: _____ Age: _____ Grade: _____

Gender: _____ Male _____ Female

Ethnicity: _____ African American/Black _____ Caucasian/White
_____ American Indian _____ Asian American
_____ Hispanic/Latino _____ Other

I consider myself (Check one):

1. _____ An excellent reader
2. _____ A good reader
3. _____ An average reader
4. _____ A poor reader

INSTRUCTIONS FOR COMPLETING THE INVENTORY

The statements listed on this inventory describes 15 strategies or actions readers use when reading academic or school-related materials such as book chapters, journal articles, stories, etc.

Directions:

Step 1: Read each statement to indicate whether you are aware of and/or use these strategies when you read.

Step 2: Use the following scale to show your strategy awareness and/or use:

1. I have never heard of this strategy before.
2. I have heard of this strategy, but I don't know what it means.
3. I have heard of this strategy, and I think I know what it means.
4. I know this strategy, and I can explain how and when to use it.
5. I know this strategy quite well, and I often use it when I read.

Step 3: After reading each strategy statement, place the numbers (1, 2, 3, 4, or 5) in the spaces preceding each statement to show your level of awareness and/or use of each strategy.

Example: _____ Sounding words out when reading

Place the number 1 in the blank space next to the strategy if you've never heard of it before; place the number 2 next to the strategy if you've heard of it, but don't know what it means; and so on.

There are no right or wrong answers to the statements in this inventory. It takes about 7-10 minutes to complete the inventory.

METACOGNITIVE AWARENESS OF READING STRATEGIES INVENTORY-REVISED (MARSİ-R, 2013)

Strategy scale:

1. I have never heard of this strategy before.
2. I have heard of this strategy, but I don't know what it means.
3. I have heard of this strategy, and I think I know what it means.
4. I know this strategy, and I can explain how and when to use it.
5. I know this strategy quite well, and I often use it when I read.

After reading each strategy statement, place the numbers (1, 2, 3, 4, or 5) in the spaces preceding each statement to show your level of awareness and/or use of each strategy.

Strategies 1-15

- _____ 01. Having a purpose in mind when I read.
- _____ 02. Taking notes while reading.
- _____ 03. Previewing the text to see what it is about before reading it.
- _____ 04. Reading aloud to help me understand what I'm reading.
- _____ 05. Checking to see if the content of the text fits my purpose for reading.
- _____ 06. Discussing what I read with others to check my understanding.
- _____ 07. Getting back on track when getting sidetracked or distracted.
- _____ 08. Underlining or circling important information in the text.
- _____ 09. Adjusting my reading pace or speed based on what I'm reading.
- _____ 10. Using reference materials such as dictionaries to support my reading.
- _____ 11. Stopping from time to time to think about what I'm reading.
- _____ 12. Using typographical aids like bold face and italics to pick out key information.
- _____ 13. Critically analyzing and evaluating the information read.
- _____ 14. Re-reading to make sure I understand what I'm reading.
- _____ 15. Guessing the meaning of unknown words or phrases.

METACOGNITIVE AWARENESS OF READING STRATEGIES INVENTORY (MARSİ-R)

SCORING & INTERPRETATION

SCORING:

The 15 strategies on the MARSİ-R inventory are scored on a 1 to 5 scale. The scores obtained provide three types information, including:

1. An individual score for each reading strategy. Review to get a cursory view of strategies you are aware of and use, and strategies you may not be aware of or use.

2. A scale score, which can be obtained by summing the items in the three reading strategy scales or categories (i.e., global reading strategies [items 1,3,5,12, & 13], problem-solving strategies [items 7,9,11,14, &15], and support reading strategies [items 2,4,6,8, &10]). To obtain scale scores, simply add up the appropriate items for each scale. Review to determine your level of awareness and use with respect to clusters or groups of reading strategies.
3. A composite score, which can be obtained by summing the scores of all strategy items in the inventory. Review to determine your level of awareness and use with respect to all reading strategies in the inventory.

Use the table below to record individual, subscales, and total inventory scores:

<i>Global reading strategies</i>	<i>Problem-solving strategies</i>	<i>Support reading strategies</i>	<i>Total reading strategies</i>
01 _____	07 _____	02 _____	GRS Mean: _____
03 _____	09 _____	04 _____	PSS Mean: _____
05 _____	11 _____	06 _____	SRS Mean: _____
12 _____	14 _____	08 _____	
13 _____	15 _____	10 _____	
GRS mean: _____ ÷ 5 = _____	PSS mean: _____ ÷ 5 = _____	SRS mean: _____ ÷ 5 = _____	Total mean: _____ ÷ 15 = _____

INTERPRETATION:

Use the following guide to interpret your scores on the MARSİ-R instrument.

1. High level of awareness (3.5 or higher).
2. Medium level of awareness (2.5-3.4).
3. Low level of awareness (2.4 or Lower).

In general, higher scores on individual, subscale, or overall reading strategies indicate higher levels of awareness and perceived use of reading strategies when reading academic or school-related materials. We recommend:

1. Using the total scores and subscale scores to derive profiles for individual students or groups of students. These profiles are useful in understanding students' levels of awareness and use of reading strategies, and in designing instruction aimed at enhancing students' awareness and use of reading strategies, which are critical for reading comprehension. For instance, lower scores on certain strategies or type of strategies may indicate a need for targeted strategy instruction based on student profile characteristics.
2. Examining the scores obtained for differences in strategy awareness and use by groups, including, but not limited to, differences between male and female students and differences between effective and struggling readers.
3. Administering the MARSİ instrument two or three times per school year to monitor growth and patterns of change in student awareness and use of reading strategies in relation to overall reading performance.