Frontline Learning Research
An official journal of EARLI

Frontline **L**earning **R**esearch Vol. 6 No. 1 (2018) 31-52
ISSN 2295-3159

A Deeper Understanding of Metacomprehension in Reading: Development of a New Multidimensional Tool

Christian Soto[a], Antonio P. Gutierrez de Blume[b], Rodrigo Asún[c], Matthew Jacovina, and Claudio Vásquez[d]

[a]University of Concepción, Chile; [b] Georgia Southern University, United States; [c]University of Chile; [d]Autonomous University of Chile

Article received 11 September 2017 / Article revised 23 February / Accepted 4 April / Available online May

Abstract

The purpose of this research endeavor was to develop and validate a new measurement tool predicated on previous research to assess learners' metacomprehension during reading. In two separate studies with Chilean undergraduate students (N = 923), we demonstrate the versatility and utility of our proposed Metacomprehension Inventory (MI). In Study 1, we provide empirical support for the psychometric soundness and construct validity of the MI. In Study 2, we provide evidence of the measurement invariance of the MI between males and females. Results of Study 1 revealed the hypothesized factor structure of the MI is sound, with high factor loadings, excellent model fit, and moderate-to-strong inter-factor correlations. Study 2 results indicated that the MI is interpreted similarly by both males and females, as factor loadings were largely statistically identical across the two groups. We discuss implications of our proposed MI for theory and applied research.

Keywords: Metacomprehension; Metacognition; Reading strategies; Factor analysis; Validity

Corresponding author: Christian Soto, Department of Spanish, University of Concepción, Victor Lamas 1290, Concepción, Chile, 4030000. E-mail: christiansoto@udec.cl DOI: 10.14786/flr.v6i1.328

1. Introduction

Many researchers propose that reading performance is improved when effective metacognitive strategies are implemented (Beck & McKeown, 1998; Beck & McKeown, 2006; Beck & McKeown, 2009; Paris & Jacobs, 1984; Sandora, Beck, & McKeown, 1999), such as selecting strategies from one's repertoire, effectively executing said strategies, and knowing when and why certain strategies do or do not apply, given task demands. This includes both the reader's ability to *understand* and to *apply* the necessary strategies during reading. The concept of metacognition has been studied since the 1970s, led by the work of Flavell and Wellman (1977), who conducted metamemory studies. Metacognition studies were then extended to the field of reading and to applied educational work (Flavell, 1979; Garner, 1987, Brown, Armbruster, & Baker, 1986; Paris & Paris, 2001, Baker, 2002; Pressley & Block, 2002; Block & Pressley, 2007; Hacker, Dunlosky & Graesser, 2009, Azevedo & Aleven, 2013), such as how to improve self-regulated learning skills in the classroom or while studying at home during learning episodes. A crucial challenge for studying metacognition is how to measure its components. Several questions arise from this challenge, particularly when studying how metacognitive knowledge, skills, and strategies influence reading comprehension. When considering *metacomprehension,* researchers must consider questions such as how metacognition influences a reader's understanding of a text, how knowledge of metacognition relates to and differs from the enactment of metacognitive strategies, and which methodologies are most appropriate. In this paper, we describe research related to these questions and propose a new metacomprehension inventory that attempts to use what has been learned about metacognition and reading to build a thorough representation of readers' metacognitive knowledge and their conscious use of strategies.

The relationship between metacognition and cognition for reading processes historically has not been very clear. A primary complexity has been determining the relative weight and importance of each component of the metacognitive process (Jacobs & Paris, 1987). A second complexity is that readers must have some awareness of their metacognitive processes in order for it to be accurately measured (Jiménez, Puente, Alvarado, & Arrebillaga, 2009). A related third complexity concerns the methods to measure some of the specific processes of metacognition (Jiménez et al., 2009), as each method has its limitations and advantages, and is aimed at examining specific aspects of metacognition (Pressley & Afflerbach, 1995). Finally, a fourth complexity has to do with the empirical evidence on the relations between metacognition and reading comprehension, which underpins the question of which methods are most appropriate for studying these processes together. The different methods to assess metacomprehension give indications of how the different components of metacognition and reading may be linked. Findings that link reading comprehension training with metacognitive measures are also useful for understanding their relations. Methods such as collecting performance/confidence judgments or asking readers to think aloud have been very useful in resolving some of these complexities and ambiguities (Bol & Hacker, 2001; Bol, Hacker, O'Shea, & Allen, 2005; Dunlosky, Rawson, & McDonald, 2002; Dunlosky, Griffin, Thiede & Wiley, 2005; Dunlosky & Lipko, 2007; Hacker, Bol, & Bahbahani, 2008; Thiede, Griffin, Wiley, & Redford, 2009; Gutierrez & Schraw, 2015). Using inventories, however, has been one of the most common methods, largely because of their practicality (e.g., Jacobs & Paris, 1987; Swanson, 1990; Pintrich & De Groot, 1990; Schmith, 1990, Schraw & Dennison, 1994; Mokhtari & Reichard, 2002). These inventories are normally short and easy to administer during different stages of reading. A single thorough metacomprehension inventory could be an excellent tool by itself or when combined with other methods to delve deeper in exploring the relation between metacognition and reading comprehension. For this reason, it is vital for inventories to evolve as new perspectives appear in the field. Accordingly, a number of tools have been developed across different languages (see Table 1).

Table 1

Metacognitive Inventories Used in Psychological Research

Name	Citation
Index of Reading Awareness (IRA)	Jacobs & Paris, 1987
Metacognitive Questionnaire (MQ)	Swanson, 1990
Motivated Strategies for Learning Questionnaire	Pintrich & De Groot, 1990
Metacomprehension Strategy Index (MSI)	Schmith, 1990
Metacognitive Awareness Inventory (MAI)	Schraw & Dennison, 1994
Reading Strategy Use (RSU)	Pereira-Laird & Deane, 1997
Metacognitive Awaraness of Reading Strategies Inventory (MARSI)	Mokhtari & Reichard, 2002
Cuestionario de Metacomprensión Lectora [a]	Peronard, Crespo, Velásquez, &Viramonte, 2002
Escala de Conciencia Lectora. (ESCOLA) [b]	Jiménez, V., Puente, A., Alvarado, J., & Arrebillaga, L. (2009)
Escala de Evaluación de la Autorregulación del Aprendizaje a partir de textos (ARATEX-R) [c]	Núñez, Amieiro, Alvarez, García & Dobarro, 2015
Revised Metacomprehension Scale (RMCS)	Zabrucky, Moore, Lin, & Cummings, 2015

Note. [a] Reading Metacomprehension Inventory; [b] Reading Awareness Scale; [c] Evaluation of Self-Regulation of Learning from Texts Scale.

Most of these inventories or questionnaires explore people's knowledge (declarative, procedural, conditional) or control/ regulation of the cognitive processes (planning, monitoring, debugging, information management, and evaluation) that are important while reading. Questions usually focus on planning, monitoring and evaluation, or ask about strategies and regulatory processes that influence comprehension. Some inventories are focused specifically on reading, whereas others are designed to be more general measures of metacognition across multiple domains. In the following paragraphs, we briefly describe some of these instruments.

One of the most popular inventories is the Metacognitive Awareness Inventory (MAI; Schraw & Dennison, 1994), which includes 52 questions that assess separate *knowledge of cognition* (declarative, procedural, and conditional knowledge) and *regulation of cognition* (planning, monitoring, debugging, information management, and evaluation) factors which are purportedly domain general. Sample items include "I try to use strategies that have worked in the past" (procedural knowledge); "I reevaluate my assumptions when I get confused" (debugging); and "I ask myself if I have considered all options after I solve a problem" (information management), answered using a 5-point Likert scale. Schraw and Dennison (1994) reported—in two separate experiments—the MAI to have a stable and consistent two-factor structure.

Another influential inventory is the Metacognitive Awareness of Reading Strategies Inventory (MARSI)**.** MARSI is a reading awareness scale, measuring metacomprehension skills in students from 6th through 12th grade. It consists of 30 items and uses a 5-point Likert scale. MARSI assesses three different

dimensions of metacomprehension: *global reading strategies*, *problem solving strategies* and *support reading strategies*. According to the authors, the first factor (global reading strategies) contains 13 items about readers' intentional strategies for analyzing a text at a global level, "setting the stage for the reading act" The second factor, *problem-solving strategies*, contains 8 items about readers' strategies for repairing problematic comprehension, particularly when a text is challenging. The third factor, *support reading strategies*, contains 9 items about strategies that readers employ that use outside reference materials, notes, or consulting others to check their understanding. Sample items include "I skim the text first by noting characteristics like length and organization" (global reading strategies); "I try to get back on track when I lose concentration" (problem solving strategies); "I discuss what I read with others to check my understanding" (support reading strategies).

Escala de Conciencia Lectora (ESCOLA; i.e., Reading Awareness Scale) is a reading awareness scale, measuring metacomprehension skills in Spanish speaking students between 8 and 13 years of age. It consists of 56 questions with three possible answers for each. ESCOLA assesses three different dimensions of metacomprehension: *planning*, *monitoring* and *evaluation*. Each question has a correct answer, and an answer that gives partial credit. Thus, the measure is designed to capture students' metacognitive competency. *Planning* questions measure readers' knowledge of how to select the most appropriate reading strategies to achieve their reading goal. *Monitoring* questions measure students' ability to adjust attention and effort during the reading task. *Evaluation* questions measure students' awareness about whether they appropriately understood the text. Sample items (translated by one of the authors of this paper) include "Before you start reading, what do you do to help in the reading process? a) I do not make any plans, just start reading [0 points], b) I consider why I'm going to read [2 points], c) I choose a comfortable place to read [1 point]" (planning); "If you are reading a book and find a paragraph difficult to understand, what do you do? a) I stop to think about the problem and how to fix it [2 points], b) I do not keep reading because I cannot solve the problem [0 points], c) I continue to read to see if the meaning is clarified later [1 point]" (monitoring); "In carrying out the activity of reading: a) I think it is useful to assess whether I understood what was written [2 points], b) I think that the evaluation is good but needs to be made by an older person [1 point], c) I do not think that after reading assessment is no longer useful [0 points]" (evaluation). ESCOLA's monitoring dimension is similar to the concept of regulation during comprehension, whereas the evaluation dimension is similar to the self-assessment of comprehension. ESCOLA has been used and validated in Spanish-speaking populations in both Spain and Argentina, and it has demonstrated sound psychometric properties (Jiménez, Puente, Alvarado, & Arrebillaga, 2009; Puente, Jiménez, & Alvarado, 2009).

Each inventory has a specific focus depending on the emphasis and framework. Some help to identify a specific component of the metacomprehension process, emphasizing a person's awareness of these processes. For example, the MAI considers very interesting distinctions between different learning processes, specifically between knowledge and control, as well as some interesting sub-mechanisms of control like debugging. However, this inventory largely focuses on learning in general and not specifically about reading comprehension. MARSI, on the other hand, proposes a particular conceptualization of the metacognitive strategies involved in reading, including global reading strategies, problem solving strategies and support reading strategies. These concepts, however, do not consider distinct time points during reading (e.g., reflection during and after reading), and some of the strategies are not commonly used by students. ESCOLA differentiates between the different time points during reading comprehension (planning, monitoring and evaluation) but the questions do not clearly situate a reader in terms of his/her current metacognitive knowledge or use of strategies. For example, in one planning question that asks what tasks a reader completes before starting to read, the possible responses are that the reader "does not make any plans," "[chooses] a comfortable place to read," and "I consider why I'm going to read," and these answers receive 1, 2, and 3 points respectively. Although it is clear that the third response is the most relevant to planning, it is less clear how choosing a comfortable place to read relates to planning. Likert-scale type questions are therefore the preferred choice due to the nature of the variables.

We agree with Schraw and Dennison (1994) who, through the MAI, make a distinction between the *knowledge of cognition* as an initial process, and the *regulation of cognition* as a subsequent process. We also agree with making distinctions between different times during the reading comprehension process. However, there is some confusion between evaluation and regulation processes, and typically their distinction is not deeply considered in metacognitive inventories, despite its importance in understanding the metacognition of reading comprehension in adult populations. This becomes an increasingly important distinction when considering how readers evaluate and regulate during different times or in different situations during the reading comprehension process.

1.1. Critical Concepts for a New Inventory

Researchers have described monitoring as including different processes. According to Hacker and his colleagues (Hacker, 1998; Keener & Hacker, 2012), monitoring has often been discussed as including both the processes of evaluation and regulation. From this perspective, readers' monitoring would be said to be successful only when they, for example, both noticed that they did not understand some part of the text and also deployed cognitive effort to remedy their understanding (e.g., by rereading the section). An alternative view of monitoring is that there is a clear distinction between the processes of monitoring (e.g., evaluating comprehension), and regulation (e.g., doing something to fix comprehension deficits) (Boekaerts, 1999; Schraw & Dennison, 1994; Schraw & Moshman, 1995). We adopt this latter view, as it is more useful in analyzing metacomprehension and its influence on comprehension. That is, failures can occur either at the monitoring or regulation stages, and both are interesting for the study of metacomprehension.

However, studies of regulation and monitoring not only have different methods and sources of research but would reflect different dimensions of the metacomprehension process. In fact, a failure to detect inconsistencies might not necessarily indicate failures in monitoring understanding, but in this case the reader might be monitoring for purposes unrelated to detecting errors (Hacker et al., 1994).

There is a close relation between monitoring and regulation, because the regulation is implemented after a reader's preliminary assessment of his/her understanding. The regulatory process acts as the reader takes action to repair or improve his/her understanding, such as by rereading a part of the text that generated confusion. In short, regulation includes adjustment operations during the comprehension process.

Evaluation can happen in different stages of reading, both during reading and after a cycle of reading. Therefore, an effective inventory should consider these different time points. On the other hand, regulation is a dynamic process that depends on evaluation, but may be implemented for different reasons, such as fixing faulty comprehension or deepening understanding. Because traditionally studies on regulation have focused on error detection when reading materials contain inconsistencies, there has been little consideration of regulation as a mechanism to understand ideas more deeply during reading. Following from this point of view, there should regularly be situations in which readers decide (consciously or subconsciously) to improve their mental representation using different strategies, even when there is no inconsistency in comprehension. As we describe next, this perspective can be incorporated in a metacomprehension inventory that can assess each one of these components separately, giving a new perspective to both theoretical and applied research, enriching the extant research between these metacognitive components and the reading comprehension measures using this type of tool (Azevedo, 2009).

1.2. Metacomprehension Inventory (MI)

The *Metacomprehension Inventory* (MI) described in this paper attempts to combine several of the strengths of the previously described inventories, while minimizing the impact of their weaknesses. Its goal is to tap into readers' evaluative and regulatory processes at different time points during reading. Throughout, the items focus on specific strategies that readers may be employing. This allows the inventory

to probe readers' conscious, strategic behaviors and does not rely on more general questions that might simply ask if readers use strategies at all. Further, the inventory uses the distinction made by Schraw and Dennison (1994) between knowledge about cognition and the control/regulation of cognition. Within knowledge of cognition, we consider the three sub-processes described by Schraw and Dennison: declarative knowledge about a reader's personal qualities and strategies in general, procedural knowledge about how to use strategies, and conditional knowledge about when and for what purpose to use strategies given task demands.

Below, we outline the different dimensions, sub-dimensions and components of the MI, that correspond to different stages of the reading process, and are measured with different items.

- *Knowledge about Cognition (KaC).* This dimension considers three components, which we assume to be strongly correlated, so that they will not form independent dimensions: a) declarative knowledge: these items refer to readers' knowledge about strategies ("I know which might be the characteristics of a good reader"), and about their knowledge of their personal reading skills ("I know the strong and weak points of my reading skills"); b) Procedural knowledge: these items refer to readers' knowledge of how to employ strategies for improved comprehension ("I know how to deal with a text to make it easier to understand for me"); Conditional knowledge: these items refer to readers' knowledge of when it is appropriate to employ comprehension strategies ("I know how to overcome a difficulty when I have problems understanding a text").

Control of Cognition (CoC): This dimension is more complex and includes three sub-dimensions which we propose exhibit some degree of statistical independence from one other:

- *Control of Cognition-Planning (CoC-P)*. These items refer to activities that occur prior to beginning a reading task such as "I question myself about the topic before starting reading." and "When I prepare myself to read, I organize my time and reading activities to finish the task on time." These items relate to control processes that a reader undergoes before beginning the task insofar as they ask whether readers strategically plan their time based on their estimates of task difficulty.
- *Control of Cognition-Evaluation (CoC-E)*. These items refer to readers' tendency and ability to examine their reading understanding. We assume that this ability subsumes two temporal sub-dimensions:
    - *Control of Cognition-Evaluation during Reading (CoC-EdR)*. These items refer to readers' tendency and ability to examine their understanding during reading (e.g., "While I'm reading I can determine how much I'm understanding").
    - *Control of Cognition-Evaluation after Reading (CoC-EaR)*. These items refer to a readers' tendency and ability to examine their understanding after completing a phase of reading (e.g., "When I finish reading a text I can know which part was more confusing to me"). These questions do not make it explicit what is meant by "after reading," and so each reader may develop his/her own interpretation (i.e., it is receptive to individual differences). The goal of these items is to gauge reflection as an anterior process as opposed to reflection in the moment (i.e., "during reading") or posterior.
- *Control of Cognition-Regulation (CoC-R)*. These items refer to readers' tendency and ability to regulate their reading understanding. We propose that this ability also subsumes two sub-dimensions:
    - *Control of Cognition-Regulation after Problematic Understanding (CoC-RPU)*. These items refer to readers' tendency and ability to engage processes and activities

to repair their understanding when they are confused, feel challenged, or notice a discrepancy (e.g., "When I find some text information strange I stop and read the paragraph more than once").

- *Control of Cognition-Regulation to Deepen Comprehension (CoC-RDC)*: These items refer to readers' tendency and ability to use reading strategies to improve their comprehension on a regular basis in an attempt to enhance their understanding (e.g., "When I read, I try to explain the text to myself using my own words"). These items explicitly do not mention any difficulties in understanding, and are meant to capture readers' strategic regulatory behaviors that are used normally when reading.

Considering readers' metacognitive knowledge and control processes separately, and at different time points, is crucial for informing interventions and building a theoretical understanding of metacomprehension. As we stated earlier, failures can occur for different reasons and at different stages, and pinpointing the nature and timing of these failures is critical for repairing readers' comprehension skills and determining appropriate remedial strategies. Theoretically, a complete model of metacomprehension should describe when strategic processes are most important; that is, a model must predict when automatic monitoring processes are unlikely to be successful or are potentially not being enacted at all. Thus, our proposed inventory, which involves proposing a fourth-order factorial model, is intended to be combined with behavioral data to inform such a model.

1.3. Research on Gender Differences in Metacognitive Monitoring

Research has shown that male and female learners may experience metacognitive monitoring in distinct ways (e.g., Ackerman, Nocera & Bargh, 2010; Denham et al., 2012; Klassen & Chiu, 2010). However, extant research on this topic has been inconclusive. Ackerman and associates (2010), for instance, found that gender was predictive of lower self-rated driving ability such that females were underconfident regarding their confidence in performance judgments in their driving ability, and that this effect remained even after controlling for baseline driving ability. Research on math achievement revealed that males not only exhibited higher achievement than females but that females were underconfident in their math achievement, and that this math performance miscalibration was more pronounced among females (Klassen & Chiu, 2010; Özsoy, 2012; Sheldrake, Mujtaba, & Reiss, 2014). On the other hand, research by Nietfeld, Shores, and Hoffman (2014) did not uncover a significant gender effect on metacognitive monitoring bias or accuracy, as males and females rated confidence similarly and exhibited near similar accuracy. Thus, to better disentangle this gender effect, we investigated whether measurement invariance exists between males and females on reading metacomprehension as an index of metacognitive monitoring.

1.4. Situating the Present Research Endeavor

Predicated on the literature review we surveyed, we adopted a two-study approach. In Study 1, our main objective was to evaluate the factorial structure of the MI. Because our model was based on substantive theoretical claims, we proposed an *a priori* hypothesized model and employed confirmatory factor analysis (CFA) techniques to assess the validity of the model. Thus, our research question for this study was: Do the observed data support our proposed latent variable model of metacomprehension in reading among a sample of Chilean undergraduates?

Hypothesis 1: We predicted that the observed data would support our hypothesized conceptualization of latent factors involved in reading metacomprehennsion in adults, as shown in Figure 1. We expected this model to fit the observed data exceptionally well.

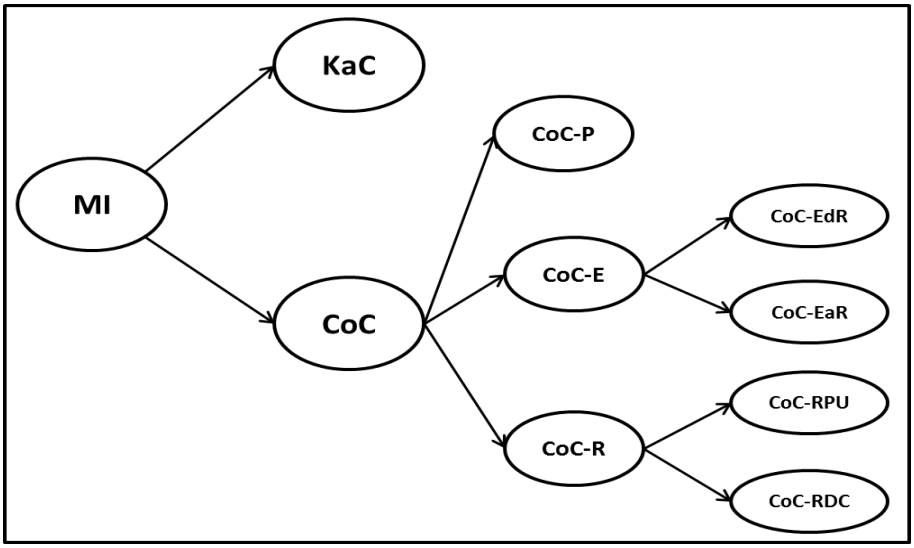

Figure 1. Hypothesized factor structure of the Metacomprehension Inventory.

In Study 2 we were interested in exploring the invariance of the model between males and females. Hence, our research question in this study was: Does the factor structure of the Metacomprehension Inventory remain consistent among a sample of male and female undergraduate students?

Hypothesis 2: Even though research shows that males and females at times vary in mean scores on metacomprehension scales, we predicted that the factor structure of the MI would remain invariant among males and females.

2. Methodology

2.1 Participants

The participants for both Studies were 923 undergraduate students of a Chilean university located in the city of Talca (Autonomous University of Chile), whose students are first generation university students. They were selected using a simple random sampling method from a population of 6,525 students. Sampling error was 3%, with a 95% confidence level assuming maximum variance.

The sample comprised 373 males, 545 females, and 5 who opted not to report gender. The average age was 22 years old (*SD* = 2.83), and participants were enrolled mainly in the disciplines of health, education, social sciences and business (see Table 2).

Table 2

Participants by Major

Department	*n*	%
Health sciences	300	32.5%
Education	245	26.5%
Social sciences	156	16.9%
Business administration	100	10.8%
Law	56	6.1%
Architecture & construction	35	3.8%
Engineering	31	3.4%
Total	923	100%

2.2 Instruments

We used a questionnaire with sociodemographic questions (i.e., gender, age, year of study and career of study) as well as the 42 Likert scale items of the MI. The MI was developed following the fourth-order factor structure shown in Figure 1. Each endogenous dimension or sub-dimension was measured on a 5-point Likert scale. All *KaC* items were answered using a response format from *strongly disagree* to *strongly agree*. In contrast, for the *CoC* dimension we used response format from *never* to *always*.

2.3. Procedure

The survey was administered face-to-face in December 2015, after obtaining approval of the directors of the different departments, Institutional Review Board (IRB) approval, and informed consent of all the participants. No student refused to participate in the study, each knowing that participation was voluntary. Students took about 20 minutes to complete the questionnaire.

2.4. Data Analysis

Two studies were conducted: in the first one we used confirmatory factor analysis (CFA) procedures and *t*-student tests to validate the MI scale by construct and convergent validity methods, whereas the second study assessed the measurement invariance between male and female students by means of the confirmatory factor model validated in the first study using multigroup CFA.

3. Results

3.1. Descriptive analysis of items

Table 3 shows descriptive statistics for each item, which allows an evaluation of some of its psychometric properties. Since all items are written such that greater values indicate a higher level of metacomprehension, no reverse coding was necessary and all means can be interpreted consistently.

Table 3

Means, SD and Skewness of the Metacomprehension Inventory Items (range of response 1 to 5)

Item	Mean	*SD*	Skewness	Item	Mean	*SD*	Skewness
1	4.0	0.8	-1.1	22	4.0	0.8	-0.3
2	4.1	0.7	-1.1	23	3.7	0.9	-0.3
3	3.7	0.9	-0.5	24	3.7	0.8	-0.3
4	3.6	0.9	-0.4	25	3.7	0.8	-0.4
5	3.6	0.8	-0.5	26	3.4	1.0	-0.3
6	3.7	0.9	-0.6	27	3.9	0.8	-0.5
7	4.1	0.7	-0.9	28	4.1	0.8	-0.4
8	4.2	0.8	-0.8	29	4.2	0.8	-0.9
9	2.9	1.1	0.1	30	3.9	0.9	-0.5
10	3.6	0.8	-0.3	31	3.9	0.9	-0.5
11	3.9	0.9	-0.5	32	4.0	0.8	-0.5
12	3.6	0.9	-0.3	33	4.0	0.8	-0.3
13	3.2	1.0	-0.2	34	3.9	0.8	-0.3
14	3.3	1.0	-0.2	35	4.1	0.8	-0.4
15	4.0	0.8	-0.4	36	4.0	0.9	-0.7
16	3.4	1.0	-0.3	37	4.0	0.9	-0.8
17	3.6	0.8	-0.2	38	3.8	0.9	-0.4
18	3.5	0.9	-0.4	39	3.9	0.9	-0.5
19	3.4	0.9	-0.4	40	3.6	1.1	-0.4
20	4.1	0.8	-0.6	41	3.6	1.0	-0.4
21	3.8	0.9	-0.5	42	3.9	0.8	-0.4
Total					3.8	0.9	-0.5

As seen in Table 3, participants generally declared relatively high levels of metacomprehension (M = 3.8). The only exception was item 9 (*I question myself about the topic before starting reading*), which had answers slightly under the midpoint of the range of responses (*Med.* = 3.0). Most of items obtained mild negative asymmetry and low kurtosis (only items 1, 2 and 7 showed asymmetry less -1 or kurtosis greater than 1). Finally, the standard deviations of the answers demonstrate that all the items were able to discriminate between the subjects.

Information presented in Table 3 does not justify removing any item for low capacity of discrimination, ceiling or floor effects, or very high asymmetry (i.e., non-normal distributions). The foregoing is confirmed by the high significant correlations between the items and the corrected total score, which fluctuated from 0.33 to 0.66.

3.2. Study 1: Validation of the Metacomprehension Inventory

3.2.1. Construct Validation

To assess the validity of the measurement model from which the items comprising the MI were generated, we evaluated several CFA models. If the theoretical model fits the observed data, this fact can be interpreted as evidence of construct validity of the instrument (Messick, 1995).

We detected that 1.2% of the test responses were missing, so, in order to verify that the missing data pattern was missing completely at random (MCAR), Little's MCAR χ^2 statistics (Little & Rubin, 1989; Schaeffer & Graham, 2002) were estimated. A significant χ^2 (i.e., $p > .05$) would suggest that the pattern of missing data is not MCAR (i.e., missing not at random [MNAR]), which poses a problem for interpretation of results because they may be biased due to systematic differences in non-responses. The result (χ^2= 1708.4, df=1695, p=.40), suggests that the missing pattern in the data was MCAR, and thus, working only with complete answers will not produce bias in results.

The CFAs were conducted on the polychoric correlation matrix of responses using Unweighted Least Squares (ULS) estimation, because the use of both allow for correcting the bias in factor loadings exhibited by the classic factor analysis based on Pearson correlations when analyzing ordinal items (Asún, Rdz-Navarro, & Alvarado, 2016; Forero, Maydeu-Olivares and Gallardo-Pujol, 2009). Statistical analysis was conducted using *Mplus 6.11* (Muthén and Muthén, 2011), which allows a pairwise deletion of missing responses.

The fourth-order factorial model shown in Figure 1 had a good fit to the data (χ^2 (N=923, df=430)=1018.04, p=.001; CFI=.919; TLI=.914, RMSEA=.057 ($CI_{90\%}$=.052, .061); SRMR=.053), and factor loading sizes ranging from .417 to .793. However, the *CoC-EdR* and *CoC-EaR* factors shared 95% of their variance, while *CoC-RPU* and *CoC-RDU* shared 90%, which does not justify considering them independent dimensions. Hence, we opted, based on this evidence, to combine these factors.

Goodness of fit of a third-order model was evaluated, merging *CoC-EdR* with *CoC-EaR* items, and *CoC-RPU* with *CoC-RDU* items. The fit of that model was good (χ^2 (N=923, df=417)=988.77, p=.001; CFI=.916; TLI=.912; RMSEA=.057 ($CI_{90\%}$=.053, .060); SRMR=.051), and very similar to the first. Also, factor loadings were high, ranging from .406 to .776. Consequently, we decided to retain this more parsimonious model in lieu of the previous more saturated model.

Additionally, with the aim of producing a shorter measure (and since a simpler model was retained), the items with the poorest fit, as evidenced by factor loadings to each factor, were eliminated. The elimination process was further supported by the collinearity of our initially proposed factor structure. The result of this process was a third-order model with 23 items, shown in the Figure 2. The goodness of fit of this model was better than the previous (χ^2 (N=923, df=389)=899.01, p=.001; CFI=.947; TLI=.940; RMSEA=.060 ($CI_{90\%}$=.057, .063); SRMR=.049), and factor loadings were higher (ranging from .445 to .791).

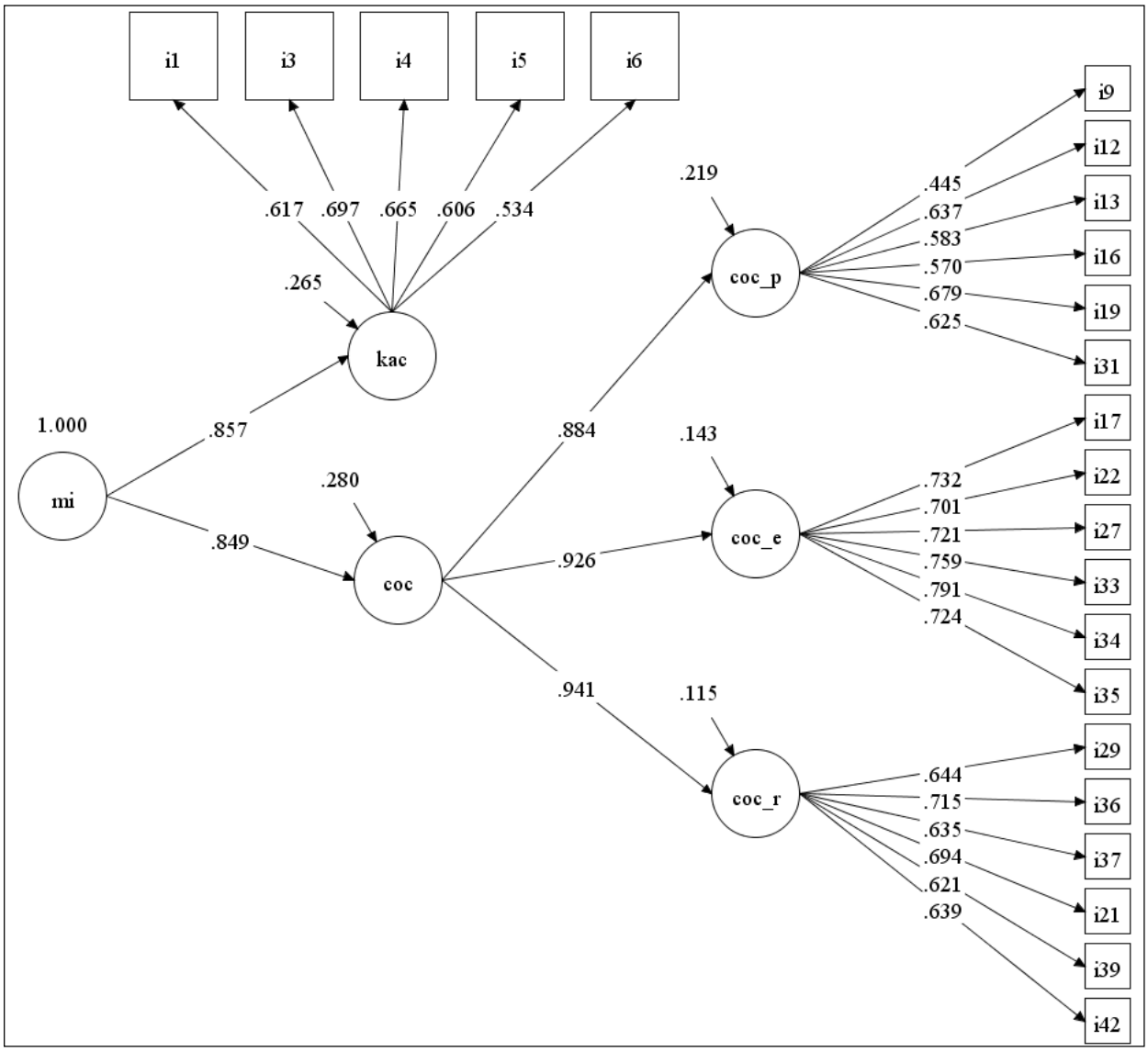

Figure 2. Final model of the Metacomprehension Inventory.

Figure 2 shows the high factor loadings of the items with their respective factors and the high relations between higher-order factors and their subordinate factors. These results provide evidence that the shorter instrument allows for global and specific scores for each latent dimension within the hierarchy of factors. Thus, we retained this best fitting factor structure as our final model.

3.2.2. Convergent Validation

To provide some evidence of the validity of the above model, we evaluated it for significant differences by gender, as there is some evidence that males and females have different levels of metacognitive monitoring (Ackerman, et al., 2010; Klaasen & Chiu, 2010; Gutierrez & Price, 2017; Sharma & Bewes, 2011), especially in countries or cultures with social norms and expectations different for men and women (e.g., Bembenutty, 2007; Bussey & Bandura, 1999). Given that our sample belongs to a Latin American country (Chile), where these different social norms still exist, we sought to examine the levels of metacomprehension between both genders.

Our analyses revealed that, compared to males, females reported greater levels of metacomprehension in all factors of the Inventory: *MI* ($t_{(916)}$ = 3.99; p<.001), *KaC* ($t_{(916)}$ = 3.61; p<.001), *CoC* ($t_{(916)}$ = 4.13; p<0.001), *CoC-P* ($t_{(916)}$ = 3.65; p<0.001), *CoC-E* ($t_{(916)}$ = 4.12; p<0.001) and *CoC-R* ($t_{(916)}$ = 4.10; p<0.001). We believe these results are evidence of the validity of the scoring of the instrument.

3.2.3. Reliability of the measurement instrument

Due to the ordinal nature of the items used, the ordinal alpha coefficient (McDonald, 1985) was employed as an indicator of the reliability (i.e., internal consistency) of the hierarchical factor structure of MI. Table 4 shows the results.

Table 4

Reliability (ordinal alpha) of the total instrument, its dimensions and subdimensions

Third-Order	Second-Order	First-Order	Reliability
Metacomprehension Inventory (MI)			.925
	Knowledge about Cognition (KaC)		.760
	Control of Cognition (CoC)		.919
		Control of Cognition: Planning (CoC-P)	.761
		Control of Cognition: Evaluation (CoC-E)	.878
		Control of Cognition: Regulation (CoC-R)	.821

N = 923

It is evident, from data in Table 4, that the MI shows high levels of reliability in both the global score and scores at all levels of the factor hierarchy.

3.3. Study 2: Invariance of the confirmatory model

In the second study, we sought to examine the measurement invariance of the validated instrument between male and female students to ascertain whether the factor structure remained consistent between the two groups, despite of their different levels of metacomprehension revealed in section 3.2.2.

3.3.1. Data Analysis

Due to the complexity of testing model invariance using statistics for ordinal data, in this study we assumed that responses were obtained on an interval scale and used maximum likelihood procedures to evaluate the models.

All data were screened for univariate and multivariate outliers according to the procedures outlined by Tabachnick and Fidell (2013) using the International Business Machine (IBM) Statistical Package for the

Social Sciences (SPSS) Statistics 22. No extreme outliers that would otherwise undermine the trustworthiness of the data were detected. The missing values analysis demonstrated that 20 cases (5.3%) in the male group and 23 cases (4.2%) in the female group had missing data. In order to verify that the missing data pattern was missing completely at random (MCAR), Little's MCAR χ^2 statistics (Little & Rubin, 1989; Schaeffer & Graham, 2002) were requested. The result of this test for the present data was non-significant for both groups (p=.53, male group and p=.18 for the female group), suggesting that the missing pattern in the data was MCAR. Thus, data analysis proceeded with 875 complete cases (353 for the male group and 522 for the female group). Furthermore, data were tested for univariate and multivariate assumptions, including multivariate normality, multicollinearity, and reproducibility of the correlation matrix via residual analysis using EQS 6.1, in order to proceed with the multi-group confirmatory factor analysis (CFA). All assumptions were met, and thus, data analysis proceeded without making any adjustments to the data.

Multi-group CFA was performed to evaluate the invariance of path coefficients among male and female groups using EQS 6.1. First, a fully constrained, fully-saturated baseline model was established for both groups to examine the feasibility of the hypothesized CFA model presented in Figure 2 by specifying the direct paths and by imposing equality constraints on all path coefficients and covariances. Subsequently, exploratory model trimming (Wald test for dropping parameters) and model building (Lagrange Multiplier [LM] test for adding parameters) procedures were interpreted in an effort to improve overall model fit of the baseline model. Next, equality constraints were individually removed for each parameter (i.e., freely estimated) that reached statistical significance at the p<.05 level using the multivariate LM χ^2 univariate increment test for releasing equality constraints. This procedure was repeated until no further parameters' LM χ^2 univariate increment reached statistical significance. This model was then deemed the final model. Releasing equality constraints for any given parameter indicates that the parameter in question differs statistically significantly across the male and female groups. Finally, the $\Delta\chi^2$ (chi-square difference) test was conducted to compare the null (i.e., fully-constrained, fully-saturated) model and the final model (i.e., released equality constraints).

3.3.2. Results

The baseline model for both groups with equality constraints imposed on all path coefficients and covariances (Figure 2) was adequately fitting to the observed data, χ^2 (447, N=875) = 777.23, p<.05, TLI=.92, CFI=.94, IFI=.94. SRMR=.05, RMSEA=.04 ($CI_{90\%}$=.03, .05). None of the model building or model trimming statistics was warranted based on theoretical considerations, and hence, this was deemed the final model for both groups.

The final model, with one statistically significant equality constraint removed, fit the observed data reasonably well, χ^2 (446, N=875) = 760.67, p<.05, TLI=.93, CFI=.94, IFI=.95. SRMR=.05, RMSEA=.04 ($CI_{90\%}$=.03, .04). The correlations between all of the metacomprehension dimensions and subdimensions were statistically significant, but did not differ significantly among the groups.

As is evident, $\Delta\chi^2$ test results between the fully-constrained baseline model and the final model with one freed equality constraint was not statistically significant, p=.89. Therefore, one can conclude that the factor structure of the MI remained consistent between males and females. The only statistically significant difference between the groups was in the path coefficient between the Control of Cognition-Regulation factor (CoC-R) and its item, *I ask myself if what I am reading is related to what I already know about the content of the text*, which was somewhat stronger among females (λ=.79) than males (λ =.70). Nevertheless, all other equality constraints, and hence, path coefficients among the groups, remained invariant. This supports our hypothesis regarding the invariance of the MI across males and females.

4. General Discussion

The purpose of Study 1 was to evaluate the validity of an innovative, more comprehensive framework for measuring metacomprehension reading strategies among a robust sample of Chilean undergraduate students (i.e., the Metacomprehension Inventory). We hypothesized that this alternative factor structure would provide a more complete and accurate representation of the latent multidimensionality of metacomprehension in reading when compared to previous conceptualizations. Our final third-order model (Figure 2) with five subordinate factors subsumed by a global metacomprehension factor demonstrated good fit to the observed data, with reasonable fit indices, low residual statistics, and factor loadings within acceptable range and in the expected theoretical direction. Given previous measures of metacomprehension (e.g., MARSI [Mokhtari & Reichard, 2002]; MQ [Swanson, 1990]; MAI [Schraw & Dennison, 1994]; ESCOLA [Jiménez et al., 2009]), the statistical evidence we provide demonstrates the more comprehensive and complete nature of our model when compared to previous models because our model assesses more fully the multidimensionality of metacomprehension in reading.

Researchers who study the latent dimensionality of measures have consistently argued that a more complete evaluation of the psychometric multidimensionality of measures that purport to assess a latent construct or set of constructs is essential for drawing more valid inferences and conclusions regarding psychological phenomena, albeit they do not always agree on how to statistically accomplish this (e.g., Chemolli & Gagné, 2014; Guay, Ratelle, Roy & Litalien, 2010; Morin, Arens, & Marsh, 2016; Morin, Arens, Tran, & Caci, 2016). Nevertheless, these researchers agree that not capturing relevant psychometric multidimensional variance by specifying simpler latent factor structures may bias goodness of fit indices, residual-based statistics, and overestimate factor correlations, possibly leading to a higher likelihood of unnecessarily inflating multicollinearity diagnostics. This leads to the potential of researchers only partially evaluating theoretical frameworks by omitting factors that are (artificially) highly correlated with other factors within the framework, and thus, providing incomplete evidence of the viability of such frameworks. Our proposed model in Study 1 mitigates these situations by proposing a more complex rather than simple model, as in previous measures of metacomprehension we surveyed. In sum, we believe that our MI will provide a more comprehensive psychometric multidimensionality of metacomprehension in reading than previous attempts.

In Study 2 we examined the measurement invariance of our proposed MI framework, as prior research has shown that metacomprehension is at times moderated by gender (e.g., Ackerman et al., 2010; Klaasen & Chiu, 2010; Gutierrez & Price, 2017; Sharma & Bewes, 2011). Even though previous research has demonstrated that self-reported metacomprehension mean scores sometimes vary among males and females, we expected the factor structure of the MI to remain invariant between the groups. Even though data of Study 1 found that there were statistically significant differences when comparing the mean scale scores between males and females, with females consistently reporting higher mean scale scores, supporting the line of research on gender differences in self-reported metacomprehension (e.g., Ackerman et al., 2010; Klaasen & Chiu, 2010; Gutierrez & Price, 2017; Sharma & Bewes, 2011), Study 2 findings support the view that the MI factor structure is consistent between males and females. Only one parameter in the factor structure of the MI was significantly different between males and females, with females showing a higher standardized path coefficient in the factor loading of the item, *I ask myself if what I am reading is related to what I already know about the content of the text*, within the *CoC-R* factor, in line with Study 1 results. However, the remaining parameter estimates remained invariant between males and females. The fact that the factor structure of the MI remained mostly consistent between this sample of Chilean adolescent males and females is encouraging because it does not bias results for either males or females, in spite of research that shows the effects of cultural and social expectations on self-reports of metacomprehension between the two groups.

4.1. Implications for Theory, Research, and Educational Practice

Theoretically this new inventory considers different metacomprehension components as with extant research, but it also adds a new perspective to approximating the mechanisms underlying metacomprehension. It deeply explores a special focus on evaluation (monitoring) and regulation. Different researchers have considered the importance of evaluation in reading comprehension (Redford, Thiede, Wiley, & Griffin, 2012), concluding that a calibrated evaluation involves awareness about when/where in the text the reader is getting a better or weaker level of understanding. However, the regulation process has enjoyed less attention in the literature, presumably because the regulation in reading comprehension represents a long tradition associated with the contradiction paradigm, and the nature of those phenomena are more complicated to assess with precision, or because some researchers prefer working with normal texts (Azevedo, 2009). However, that situation does not necessarily mean we must stop our efforts to generate new, improved approximations about the particular mechanisms underlying those phenomena, and particularly between the relation between evaluation and regulation (Keener & Hacker, 2012). This inventory has the potential to offer a new, more specific theoretical perspective because it not only distinguishes between evaluation during and after reading, but also conceptualizes self-regulation more comprehensively than previous attempts. An important contribution is that it considers regulation after difficulties in understanding, where some process is activated to evaluate and implement adjustment to repair the incoherence in the mental representation and other mechanisms within regulation to deepen comprehension. This latter process does not stem from a problem with confusion or a discrepancy, but instead, according to a typical reading process individuals sometimes employ when regulating their reading comprehension to acquire a better mental representation of the text. In fact, under this perspective to use reading comprehension strategies during reading to improve current comprehension could be considered a regulation behavior, regardless of whether the use of strategies is conscious, subconscious, controlled or automatized, while it is generated from the evaluation state to trigger the repair.

Another contribution of this work is that it better highlights the dynamic relation between cognition and metacognition, and how the relation between those two levels is more flexible than has been previously assumed, albeit a deeper discussion of this point is beyond the scope of this paper. Regarding different ways of regulation, the MI considers two different mechanisms about correcting errors in reading, capturing a more complete understanding about the metacomprehension process when using the Inventory.

4.2. Avenues for Future Research

Despite the fact that our studies used a robust sample size, they may not necessarily generalize to other populations or samples of adult learners from other cultures. Future research should verify our findings using children and adults from other cultures to ascertain whether our MI factor structure and invariance generalizes to other samples, populations, developmental stages, and cultures. Moreover, it would be worthwhile for researchers to explore the utility of the MI in actual classroom settings. Finally, future research should examine the relation of the MI to previous measures of reading comprehension and investigate whether the MI's more comprehensive nature predicts additional variance not attributable to previous measures.

4.3. Methodological Reflections and Limitations

No research endeavor involving human beings is ever without limitations. Even though the MI represents a more robust psychometric multidimensionality attempt with respect to metacomprehension, with strong support for our hypothesized higher-order factor structure, it still relies on self-report responses from participants. As has been consistently demonstrated, individuals may not be the best raters of their own attitudes, beliefs, and perceptions due to such phenomena as the social desirability bias. Moreover, our research design was cross-sectional and correlational in nature, thereby limiting the inferences and

conclusions we can draw from these data. A longitudinal design may have permitted us to investigate how enduring these perceptions are among this sample of Chilean learners. In addition, we acknowledge the common method variance dilemma that may bias correlations and effect sizes due to our single method, single rater approach. In spite of these limitations, our combined studies contribute to a deeper understanding of the complex dynamics involved in metacomprehension of reading.

4.4. Conclusion

Study 1 provided support for our more comprehensive framework for conceptualizing metacomprehension than prior research. Our MI measure is also compact, and thus, it can be readily employed with students in authentic, ecologically valid learning environments such as classrooms. Information gleaned from the MI could assist classroom teachers to uncover specific deficits in learners' reading comprehension strategy use. This would then enable teachers to develop individualized reading interventions that could benefit poor reading comprehenders in particular. Study 2 showed that, with the exception of one path coefficient that was stronger among females than males, all other parameters in the model were invariant between males and females. This is important information to know because the MI could be applied equally as well to male and female learners without the need to make adjustments. Finally, this inventory demonstrated sound psychometric characteristics and conceptual foundations that lead us to conclude that it is an excellent resource for researchers and teachers.

Acknowledgments

We thank the Autonomous University of Chile at Talca, for providing access to students for this research endeavor. A special thanks to the Centro de Estudios y Gestión Social (CEGES) for making the fieldwork involved in this research possible. Finally, we acknowledge the contribution of the IT16I10044 Project: Technology for the improvement of reading comprehension in students of the Chilean school system (Tecnología para el mejoramiento de la comprensión lectora en estudiantes del sistema escolar chileno; FONDEF, CONICYT).

Appendix: Metacomprehension Inventory

Item	Spanish (Original) Version	English Translation
1	Conozco los puntos fuertes y débiles de mis habilidades lectoras	I know the strong and weak points of my reading skills
2	Sé qué tipo de textos podrían ser más difíciles para mí	I know which topics are more complex to read for me
3	Sé cómo enfrentar un texto para que se me haga más fácil de entender	I know how to deal with a text to make it easier to understand for me
4	Conozco las estrategias necesarias para leer mejor	I know the strategies needed to read better
5	Sé cómo sobreponerme cuando tengo una dificultad al comprender un texto	I know how to overcome a difficulty when I have problems understanding a text

6	Sé cuáles podrían ser las características de un buen lector	I know which might be the characteristics of a good reader
7	Sé los temas que para mí podrían ser más complejos de leer	I know which subject-topics might be more difficult to read for me
8	Cuando algo se me hace difícil a veces vuelvo atrás en la lectura	When something is difficult to understand, sometimes I turn back into my reading
9	Me hago preguntas sobre el tema antes de empezar a leer	I question myself about the topic before starting reading
10	Después de leer un texto puedo saber con precisión el nivel de comprensión que he alcanzado	When finishing my readings I know precisely which level of understanding I achieved
11	Durante la lectura me pregunto si voy entendiendo bien o no	While I am reading I wonder whether I'm understanding the text correctly or not
12	Pienso en lo que realmente necesito comprender antes de empezar una tarea de lectura	I think about what I really need to understand before starting a reading task
13	Me propongo objetivos específicos antes de empezar una tarea de lectura	I define specific objectives for myself before starting a reading task
14	Cuando estoy leyendo me pregunto constantemente si estoy alcanzando mis metas de lectura	When I'm reading, constantly I ask myself whether I'm reaching my reading goals or not
15	Cuando estoy leyendo, de vez en cuando pienso si estoy entendiendo lo que estoy leyendo	When I'm reading, occasionally I think whether I'm understanding what I'm reading
16	Cuando me dispongo a leer organizo el tiempo y las actividades de lectura para poder terminar la tarea a tiempo	When I prepare myself to read I organize my time and reading activities to finish the task on time
17	A medida que voy leyendo puedo ser preciso/a para determinar cuánto voy comprendiendo	While I'm reading I can determine how much I'm understanding
18	Cuando respondo un test de comprensión lectora sé cómo me ha ido	When I respond to a reading comprehension test I know how good or bad was my performance
19	Pienso en las distintas maneras de abordar la lectura de un texto y escojo la mejor	I think about the different ways to accomplish a reading task and choose the best approach
20	Cuando no logro entender una parte de un texto trato de ir más despacio para entender	When I don't understand a part of a text I try to slow down my reading to understand it

21	Trato de vincular diferentes partes del texto para que tenga más sentido	I try to link different parts of the text to make more sense
22	Durante la lectura sé si estoy comprendiendo bien, regular o mal	During my reading I know whether I am understanding well, not so well or bad
23	Cuando termino una lectura puedo saber si logré las metas que me había propuesto	When I finish my reading I can tell myself whether I achieved my goals or not
24	Cuando estoy confundido/a me pregunto si lo que estaba suponiendo acerca del texto era correcto o no	When I feel confused I wonder whether my assumptions about text were correct or not
25	Cuando leo trato de estar consciente del nivel de comprensión que voy alcanzando	When I read I try to be aware of the level of understanding I'm reaching
26	Organizo el tiempo de lectura para lograr mejor mis objetivos	I organize my reading time for better achievement of my goals
27	Cuando estoy leyendo puedo determinar si alguna parte del texto está siendo más fácil o más difícil para mí	When I'm reading I can determine whether part of the text is being easier or harder to understand
28	Después de leer un texto puedo determinar qué tan complejo ha sido para mí	After reading a text I can determine how complex it has been for me
29	Cuando me resulta extraña la información del texto me detengo y la leo más de una vez	When I find some text information strange I stop and read the paragraph more than once
30	Puedo pensar en ejemplos para poder entender mejor la información del texto	I can think about examples to better understand text information
31	Leo cuidadosamente los enunciados antes de empezar con la lectura	I read carefully the instructions before starting my reading
32	Cuando estoy leyendo puedo darme cuenta si el texto es más o menos complejo para mí	When I'm reading I can tell whether the text is more or less complex for me
33	Cuando termino de leer puedo evaluar si he comprendido bien o no	When I finished reading I can evaluate whether I understood the text well or not
34	Luego de leer un texto puedo evaluar si interpreté correctamente el texto	After reading a text I can assess whether I interpreted the text correctly
35	Cuando termino de leer un texto puedo saber qué parte fue más confusa para mí	When I finish reading a text I can know which part was more confusing to me
36	Cuando el texto se me hace difícil de comprender aumento mi atención y esfuerzo	When a text is difficult to understand I increase my attention and effort
37	Cuando una palabra es extraña para mi trato de entender por el contexto de la lectura	When a word is strange to me I try to understand it by the context

38	Si no estoy comprendiendo una parte, a veces sigo con la lectura en busca de clarificación	If I'm not understanding a part of a text, sometimes I keep reading seeking for clarification
39	Cuando leo trato de ir explicándome el texto con mis propias palabras	When I read I try to explain the text to myself using my own words
40	Mientras leo trato de imaginar lo que vendrá a continuación en el texto	while I read I try to imagine what come next in the text
41	Busco hacerme preguntas para darle más sentido al contenido del texto	I seek inquire myself about what I read to give more meaning to the text content
42	Me pregunto si lo que estoy leyendo está relacionado con lo que ya sé del contenido del texto	I ask myself whether what I am reading is related to what I know about the text content

Note: In bold, item retain in final Metacognitive Inventory.

Keypoints

- The Metacomprehension Inventory (MI) exhibited sound psychometric properties.
- The MI is an effective tool for measuring reading comprehension among adults.
- The MI parameter estimates were invariant across males and females.

References

Ackerman, J., Nocera, C., & Bargh, J. (2010). Incidental Haptic Sensations Influence Social Judgments and Decisions. *Science, 328,* 1712-1715. doi: 10.1126/science.1189993

Asún, R. A., Rdz-Navarro, K., & Alvarado, J. M. (2016).Developing multidimensional Likert scales using item factor analysis: The case of four-point items. *Sociological Methods and Research, 45*(1), 109-133. doi: 10.1177/0049124114566716

Azevedo, R., & Aleven, V. (2013). *International handbook of metacognition and learning technologies*. New York: Springer.

Azevedo, R. (2009). Theoretical, methodological and analytical challenges in the research on metacognition and self-regulation: A commentary. *Metacognition & Learning, 4*, 87-95. doi: 10.1007/s11409-009-9035-7

Baker, L. (2002). Metacognition in comprehension instruction. In Block, C. C. & Pressley, M. (Eds.), *Comprehension Instruction: Research-based best Practices* (pp. 77-95). New York: The Guil-ford Press.

Beck, I.L. & McKeown, M.G. (1998). Comprehension: The sine qua non of reading. In S. Patton & M. Holmes (Eds.), *The keys to literacy* (pp. 40-52). Washington, DC: Council for Basic Education.

Beck, I.L. & McKeown, M.G. (2006). *Improving comprehension with questioning the author: A fresh and expanded view of a powerful approach*. New York, N.Y: Scholastic.

Beck, I.L. & McKeown, M.G. (2009). The role of metacognition in understanding and supporting reading comprehension. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition*

in Education. The Educational Psychology Series (pp. 7-25). Mahwah, NJ: Lawrence Erlbaum Associates.

Bembenutty, H. (2007). Self-regulation of learning and academic delay of gratification: Gender and ethnic differences among college students. *Journal of Advanced Academics, 18*(4), 586-616.

Block, C. C., & Pressley, M. (2007) Best practices in teaching comprehension. In L.B. Gambrell, L. M. Morrow, & M. Pressley (Eds.), *Best practices in literacy instruction* (pp. 220–242). New York: Guilford.

Boekaerts, M. (1999) Self-regulated learning: where we are today. *International Journal of Educational Research 31*(6), 445-457. doi: 10.1016/S0883-0355(99)00014-2

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education, 69*(2), 133-151. doi: 10.1080/00220970109600653

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education, 73*(4), 269-290. doi: 10.3200/JEXE.73.4.269-290

Brown, A. L., Armbruster, B. B. & Baker, L. (1986). The role of metacognition in reading and studying. In Orasanu, J. (Ed.), *Reading Comprehension: From Research to Practice*. Hillsdale, NJ: Lawrence Erlbaum.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*(4), 676-713.

Chemolli, E. & Gagné, M. (2014). Evidence against the continuum structure underlying motivation measures derived from self-determination theory. *Psychological Assessment, 26*(2), 575-585. doi: 10.1037/a0036212

Denham S. A., Warren-Khot H. K., Bassett H. H., Wyatt T., Perna A. (2012). Factor structure of self-regulation in preschoolers: testing models of a field-based assessment for predicting early school readiness. *Journal of Experimental Child Psychology, 111*(3), 386-404. doi: 10.1016/j.jecp.2011.10.002

Dunlosky, J., Rawson, K.A., & McDonald (2002). Influence of practice tests on the accuracy of predicting memory performance for paired associates, sentences, and text material. In T.J. Perfect & B.L. Schwartz (Eds.), *Applied Metacognition* (pp. 68-92). New York, NY: Cambridge University Press.

Dunlosky, J., Griffin, T., Thiede, K. & Wiley, J. (2005). Understanding the delayed keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31*, 1267-1280. doi: 10.1037/0278-7393.31.6.1267

Dunlosky, J., & Lipko, A.R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*(4), 228-232. doi: 10.1111/j.1467-8721.2007.00509.x

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906-911.

Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail y J. W. Hagen (Eds.), *Perspectives on the Development of Memory and Cognition*. Hillsdale, N.J.: Erlbaum.

Forero, Maydeu-Olivares and Gallardo-Pujol, (2009). Factor Analysis with Ordinal Indicators: A Monte Carlo study DWLS and ULS estimation. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 625-641. doi: 10.1080/10705510903203573

Guay, F., Ratelle, C., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences, 20* (6), 644-653. doi: 10.1016/j.lindif.2010.08.001

Gutierrez, A. P., & Price, A. F. (2017). Calibration between undergraduate students' prediction of and actual performance: The role of gender and performance attributions. *The Journal of Experimental Education,85*, 486-500. doi: 10.1080/00220973.2016.1180278

Gutierrez, A. P., & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration. *The Journal of Experimental Education, 83*(3) 386-404. doi: 10.1080/2331186X.2017.1314652

Hacker, D., Plumb, C., Butterfield, E., Quathamer, D., & Heineken, E. (1994). Text revision: Detection and corrections of errors. *Journal of Educational Psychology, 86*, 65-78.

Hacker, D. (1998). Definitions and empirical foundations. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice*. Mahwah, NJ: Erlbaum.

Hacker, D., Bol, L., & Bahbahani, K. (2008). Explaining calibration in classroom context. The effects of incentives, reflection, and attributional style. *Metacognition and Learning, 3,* 101-121. doi: 10.1080/2331186X.2017.1314652

Hacker D. J., Dunlosky J., Graesser A. C. (2009), *Handbook of metacognition in education*. New York: Routledge.

Jacobs, J. E., and Paris, S. G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist. 22* (3&4) 255–278.

Jiménez, V., Puente, A., Alvarado, J., & Arrebillaga, L. (2009) Medición de estrategias metacognitivas mediante la Escala de Conciencia Lectora: ESCOLA. *Electronic Journal of Research in Educational Psychology,7*(2), 779-804.

Keener, M. C., & Hacker D. J. (2012). Comprehension monitoring. In N. M. Norbert (Ed.), *Encyclopedia of the sciences of learning*. New York, NY: Springer.

Klaasen, R., & Chiu, M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology, 102*(3), 741-756. doi: 10.1037/a0019237

Little, R.J. & Rubin, D.B. (1989). The analysis of social science data with missing values. *Sociological Methods and Research*, *18* (2&3), 292-326.

McDonald, R.P. (1985). *Factor analysis and related methods*. Hillsdale NJ: Erlbaum.

Mokhtari, K., & Reichard, C. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*(2), 249-259. doi: 10.1037/0022-0663.94.2.249

Morin, A., Arens, K., & Marsh, H. (2016). A Bifactor Exploratory Structural Equation Modeling Framework for the Identification of Distinct Sources of Construct-Relevant Psychometric Multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal 23*(1), 116-139. doi: 10.1080/10705511.2014.961800

Morin, A., Arens, K., Tran, A. & Caci, H. (2016) Exploring sources of construct-relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the Composite Scale of Morningness. *International Journal of Methods in Psychiatric Research 25*(4), 277-288. doi: 10.1002/mpr.1485

Muthen, L. K., & Muthen, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthen & Muthen.

Nietfeld, J., Shores, L., & Hoffman, K. (2014). Self-regulation and gender within a game-based learning environment, *Journal of Educational Psychology, 106*(4), 961-973. doi:10.1037/a0037116

Özsoy, G. (2012). Investigation of fifth grade students' mathematical calibration skills. *Educational Sciences: Theory and Practice, 12*(2), 1190–1194.

Paris, S. G. & Jacobs, J. E. (1984). The benefits of informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology, 76* (6), 1239-1252.

Paris, S.G. & Paris, A.H. (2001). Classroom Applications of Research on Self-Regulated Learning. *Educational Psychologist, 36*(2), 89-101. doi: 10.1207/S15326985EP3602_4

Pintrich P. & De Groot, E. (1990). Motivational and Self-Regulated Learning Components of Classroom Academic Performance. *Journal of Educational Psychology 82* (1), 33-40.

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

Pressley, M., & Block, C. C. (2002). Summing up: What comprehension instruction could be. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices*. New York: Guilford Press.

Puente, A., Jiménez, V. & Alvarado, J.M. (2009) *Escala de conciencia lectora (ESCOLA). Evaluación e intervención psicoeducativa de procesos y variables metacognitivas durante la lectura*. Madrid: EOS.

Redford, J., Thiede, K., Wiley, J. & Griffin, T. (2012).Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction, 22* (4), 262-270. doi: doi:10.1016/j.learninstruc.2011.10.007

Sandora, C., Beck, I., & McKeown, M. (1999). A comparison of two discussion strategies on students' comprehension and interpretation of complex literature. *Journal of Reading Psychology, 20* (3), 177-212.

Schaeffer, J. & Graham, J. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods, 7* (2), 147-177.

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics, 11* (2), 129-158.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19* (4), 460-475.

Schraw, G., & Moshman, D. (1995). *Metacognitive theories*. *Educational Psychology Review*, *7* (4), 351-371.

Sharma, M. D & Bewes, J. (2011). *Self-monitoring: confidence, academic achievement and gender differences in physics*. *Journal of Learning Design, 4*, (3), 1-13.

Sheldrake, R., Mujtaba, T., & Reiss, M. (2014). Calibration of self-evaluations of mathematical ability for students in England aged 13 and 15, and their intentions to study non-compulsory mathematics after age 16. *International Journal of Educational Research*, *64*, 49-61. doi: 10.1016/j.ijer.2013.10.008

Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology, 82* (2), 306-314.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.

Thiede, K., Griffin, T., Wiley, J. & Redford, J. (2009). Metacognitive Monitoring During and After Reading. In J. Dunlosky, A. Graesser & J. Hacker (Eds.), *Handbook of Metacognition in Education*. New York: Routledge.