

The unidimensionality and overestimation of metacognitive awareness in children: validating the CATOM

P. C. Ferreira, A.M. Veiga Simão and A. Lopes da Silva

Faculty of Psychology, University of Lisbon, Lisbon (Portugal).

Título: Unidimensionalidad y sobreestimación de la conciencia metacognitiva en niños: Validación del CATOM.

Resumen. Los niños normalmente tienen dificultades para informar sobre su funcionamiento metacognitivo, hecho que, con frecuencia, les lleva a sobrevalorarse en situaciones de aprendizaje. Por tanto, este estudio pretende comprender cómo los niños ($n = 1029$) informan sobre su funcionamiento metacognitivo y ofrece una primera aproximación a la medición de la conciencia metacognitiva (CM) en niños. Así, tras observar que los resultados de un análisis factorial exploratorio indican que el instrumento de medida presenta una estructura unidimensional, se aplicó la Teoría de Respuesta al Ítem para analizar dicha unidimensionalidad así como las interacciones entre los participantes y los ítems. Los resultados indican una buena fiabilidad tanto ítem (.87) como persona (.87), con un alfa de Cronbach .95 para la dimensión CM. Además, los resultados corroboran la tendencia de los niños a sobreestimar su funcionamiento metacognitivo y sugieren que el instrumento tiene un alto potencial tanto para la investigación como para la práctica profesional.

Palabras clave: metacognición, conocimiento metacognitivo, habilidades metacognitivas, aprendizaje autoregulado, Teoría de Respuesta al Ítem.

Abstract. Children often have difficulty in reporting their metacognitive functioning, which leads them to frequently overrating themselves under learning situations. Hence, this study presents a preliminary approach of how children's metacognitive awareness (MA) can be measured. Essentially, this study aims to understand how children ($n = 1029$) report their metacognitive functioning. In a first analysis, EFA revealed a unidimensional structure of the instrument (MK and MS). Item Response Theory was then used to analyse the unidimensionality of the dimension and the interactions between participants and items. Results revealed good item reliability (.87) and good person reliability (.87) with good Cronbach's α for MA (.95). These results show the potential of the instrument, as well as a tendency of children to overrate their metacognitive functioning. Implications for researchers and practitioners are discussed.

Key words: Metacognition, Metacognitive Knowledge, Metacognitive Skills, Self-regulated Learning, Item Response Theory.

Introduction

The literature on self-regulation (Pintrich, 2000) considers metacognition as one of its important components because it consists of individuals' knowledge of their own cognitive and affective processes, including their ability to consciously and intentionally monitor and regulate these processes (Hacker, 1998). Flavell (1976) first defined metacognition as active monitoring and subsequent regulation and management of processed information regarding concrete goals or objectives. Later, Efklides (2008) described metacognition as being a "critical component of the self-regulation process because" (p. 283) it includes self-awareness, which in turn, involves past experiences, beliefs and goals, as well as future goals when students think, feel and act in context.

Efklides (2011) distinguishes between three different metacognitive facets in the Metacognitive and Affective Model of SRL (MASRL) which are related to motivational and affective aspects, namely, metacognitive knowledge (MK) and metacognitive skills (MS) at the person level of SRL, and metacognitive experiences (ME) at the person-task level of SRL. While MK pertains to beliefs, declarative knowledge, theories about goals, strategies, cognitive functions, tasks and persons (Efklides, 2001), MS encompasses procedural knowledge and strategies, including planning, self-monitoring and evaluating (Veenman & Elshout, 1999). ME are described as being overt processes of cognitive mon-

itoring during the completion of a task (Efklides, 2006). These three facets comprise our operational definition of metacognition in this study.

Beliefs about ability have an impact on how individuals approach a task (Dweck, 1999). That is, the manner with which individuals view their accomplishments and failures influence their approach to new challenges. Hence, if children are to mature in life by reflecting on the decisions they make in their surrounding environments, then we feel that it is essential for them to develop metacognitive awareness (MA). In order to do so, it is essential that research focus on how MA originates and develop.

The literature (Wigfield, Klauda, & Cambria, 2011) has indicated a lack of both studies and instruments with primary school children regarding metacognitive and motivational aspects of self-regulation. Furthermore, it is still unclear and more empirical evidence is needed on how children acquire MA or specifically, MK, considering it is related to other metacognitive facets, such as MS and ME (Efklides, 2011) which will lead to the development of new MS (Misailidi, 2010). Hence, in an attempt to contribute to the literature on metacognitive functioning, this paper presents a study that proposes an approach of how the person level (MK and MS) of the MASRL model can be measured in fourth-grade children. Specifically, this study aims to understand the accuracy with which young children report their metacognitive functioning. We consider young children from infancy to the age of 11, as indicated by other authors (Larkin, 2010). Therefore, in order to achieve our objective, we chose to use Item Response Theory, which would allow us to calibrate our participants and items on a common scale (DeMars, 2010; Embretson, 1996). This type of measurement provides an analy-

*** Dirección para correspondencia [Correspondence address]:**

Paula da Costa Ferreira. Faculty of Psychology. University of Lisbon. Alameda da Universidade 1649-013 Lisbon (Portugal).
E-mail: paula.costa.ferreira@gmail.com

sis of the interactions between people and items, which would help us interpret the variables we wanted to measure. Furthermore, the interpretations of items in which participants have a higher probability of dominating, have a greater diagnostic convenience for our study than group-related ratings.

We first present other studies that discuss children's awareness of their metacognitive functioning, as well as the accuracy with which they report it, with the purpose of sustaining our target population. Then, we demonstrate how we developed and tested the Children's Awareness Tool of Metacognition (CATOM) with exploratory factor analysis and the Item Response Theory in order to help us better understand how children report their MK and MS.

Evidence of Children's Metacognitive Awareness

Evidence has shown that metacognitive abilities seem to progress with age (Kuhn & Dean, 2004; Schneider, 2008; Schneider & Lockl, 2002; Schraw & Moshman, 1995). Specifically, Schraw and Moshman (1995) proposed that children as young as age 6 develop cognitive knowledge and are able to reflect on about their cognition. Around early middle childhood, children seem to gain a considerable understanding of how the mind processes information actively through interpretation and construction and, consolidate these skills between the ages of 8 to 10 (e.g. Barquero, Robinson, & Thomas, 2003). What's more, at this age, children realize that perceptual information must be adequate and present in order to produce knowledge (Flavell, 2004). Bares (2011) for instance, suggested that children between the ages of 8 and 10 are able to use metacognitive processes on a consistent and mature basis. With time, children develop their ability to regulate cognition and seem to improve their monitoring and regulation skills by practicing planning between the ages of 10 to 14. Eventually, monitoring and evaluation of cognition may or may not develop with substantial improvements later on in life, along with the construction of metacognitive theories (Schraw & Moshman, 1995).

The literature on metacognition has provided evidence that children in primary school possess not only declarative knowledge regarding their metacognitive functioning, but procedural knowledge as well - MS (Annevirta & Vauras, 2006). Efklides (2011) proposed that metacognition could interact with self-regulation of behavior and motivational aspects at the person level (MK), including learners' beliefs about themselves and the task, and at the person-task level (ME), when the learner is engaged in the task. Essentially, the author explains how ME affects self processes and causal attributions by providing feedback about one's self and the task at hand, which ultimately, will affect individuals' awareness of themselves as learners (MK). Essentially, the author presented metacognition as being deliberate and as encompassing various strategies, which are also involved in self-regulation processes - namely, orientation strategies, planning strategies, regulation strategies of cognitive processing, moni-

toring strategies, evaluation strategies and recap strategies. These strategies may be initially used by children unconsciously, although they gain an awareness of this use with time. Eventually, children learn to use these strategies intentionally in a self-regulated way (Pihlainen-Bednarik & Keinonen, 2010; Schneider & Lockl, 2002).

Thomas and Au Kin Mee (2005) discovered how primary school children were familiar with the names of the strategies they used, how they used them and how they could be beneficial to them while they learned. The authors presented evidence regarding students awareness of the strategies they used due to the development of metacognition. In general, the literature on metacognition shows that students who are more effective at regulating their cognitive strategy use, also demonstrate more adaptive performance and achievement outcomes (Baker, 1994; Butler & Winne, 1995; Schraw & Moshman, 1995).

Schneider (2008) for instance, investigated the relationship between theory of mind at age 3 and the subsequent development of metamemory at age 5 with 174 children. Essentially, ToM pertains to the "ability to estimate mental states, such as beliefs, desires, or intentions, and to predict other people's performance based on judgments of their mental states" (p. 115). Schneider theorized that theory of mind enabled young children to acquire MK and language skills more easily, and argued that developing early theory of mind competencies could facilitate the development of metamemory later on. Specifically, the results of this study revealed that while MK had a tendency to increase with age, MS were not so evident. We mention theory of mind in our study because we agree with Flavell (2002) that it encompasses pretty much the same objective as metacognition, which is to study children's knowledge and cognitive development about what goes on in their minds as they learn.

On another note, Burman (1994) cautions that developmental psychology cannot be considered an absolute scientific doctrine with normative standards by which children must be compared to. In this sense, the author advises that general standardization of children's development through general measuring be avoided because of the complexity surrounding these children's learning and living environments. Furthermore, some evidence has revealed that general metacognition does not inevitably enhance with age. As an example, Sperling, Howard, Miller and Murphy (2002) measured general metacognitive knowledge and regulation in children from grades 3 to 8 with a validated self-report measure. The authors discovered that younger students had higher metacognition scores than older students. What's more, the authors hypothesized that because the instrument they applied measured general metacognition, that metacognition could possibly be more domain-specific as students become older and attain more expert content knowledge.

Larkin (2010) suggested that engaging young children in experiences which facilitate metacognitive development, encourages learners to be responsible for their own learning and to interact with others in meaningful ways. Furthermore,

there is a need to use MK and MS in specific subject areas in the sense that metacognition is transversal, but specific to each area. The author refers to the instruction of the English language, which specifically includes metalinguistic knowledge, such as various parts of speech (such as nouns, verbs, adjectives etc.), as well as morphological and phonological aspects, and language style and tone. In addition, Larkin explains how over time and with experience and instruction, children are able to develop this specific type of MK, which will allow them to differentiate between particularities of the language, such as letters, sounds and meaning. Later, Kirsch (2012) also focused on metacognition in a language learning context and demonstrated how MK was essential for children learning a foreign language to develop self-regulation, autonomy and proficiency. Correspondingly, this study focuses on children's MA (MK and MS) when learning English as a foreign language (EFL).

Children's Accuracy in Reporting Metacognitive Awareness

Some of the literature has suggested that young children are less accurate than older children at predicting how well they will be able to learn something (ease of learning judgments), as well as judging how well they have learned something (judgments of learning). These metacognitive judgments are contemplated in the MASRL model proposed by Efklides (2011) and seem to be more accurately produced by children throughout the elementary school years (Schneider, 1998). Some authors argue (Rizzo, Steinhausen, & Drechsler, 2010) that children in this age group are capable of making accurate and differential judgments of their self-regulation processes and hence, be metacognitively active. Others posit that both children (from 8 years of age on) and adults are weak at determining good from bad performance because of their inaccurate confidence judgments (Allwood, Ask, & Granhag, 2005; Allwood, Innes-Ker, & Fredin, 2008).

Flavell, Friedrichs and Hoyt (1970) studied children's prediction accuracy using a performance prediction paradigm. Essentially, the authors asked a sample of nursery school children, kindergarteners, second-grade and fourth-grade students to predict how many pictures (from 0 to 10) they could remember. In order to conduct this task, children were presented with a new picture every new trial. Although the nursery school children and kindergarteners were more overconfident than the second and fourth grade, children, all of the children's predicted memory span was higher than their actual memory span. Similarly, Shin, Bjorklund and Beck (2007) asked kindergarteners, first-graders, and third-graders to predict the numbers of pictures they could remember out of 15 in a supraspan task. As in Flavell et al's study (1970), the younger children were more overconfident and overestimated more than the third-graders. Hence, the authors stated that when children think they are better than what they actually are on a specific task, leads them to having higher levels of motivation to persist on that task, which may

result into better performance in comparison with more accurate children. This is consistent with Bandura's theory on self-efficacy (1989), considering children may benefit from overestimating their performance because they continue to be motivated on a particular task.

Lipko-Speed (2013) found similar results, but mentioned that children's overestimation perseveres even with practice due to the lack of knowledge transfer. In short, the past can predict the future in terms of performance. Furthermore, this constant overestimation may lead to continuous failure in certain tasks when confronted with feedback. This is especially true if these children believe that the amount of effort they make alone absolutely translates into a successful performance on a task (Stipek & MacIver, 1989). Specifically, if children believe that effort, rather than knowledge regarding their previous performance, is a good indicator of their future performance, than it is probable that they will continue to be overconfident and may not pursue improvements in their performance. Essentially, children may not adjust their behavior in order to enhance task performance and may even avoid asking for help from teachers, colleagues, or parents.

Thus, in light of the theoretical findings and recommendations we have presented in this section, we want to develop a new measure and to understand how children view themselves as metacognitively active agents of their learning process in their English as a foreign language (EFL) class. Specifically, we want to know how children report their MA (MK and MS) in EFL and hence propose that:

H1: Children overrate their MA in EFL classes.

Method

Participants and Learning Context

A total of 1029 students participated in this study. Specifically, our sample consisted of 23 students in the development of the items of the CATOM, 805 students (mean age = 8.85; $SD = .70$; 50.2% boys) in the exploratory factor analysis (EFA) and 201 students (mean age = 9.37; $SD = .52$; 50.0% boys) in the IRT analysis. All students were in the fourth-grade, had the same level of English according to the Common European Reference for Language Learning (level A1) and were from 9 different schools in the district of Lisbon. The children that participated in this study were predominantly of Portuguese nationality (86%). Other students were of different origins (i.e., African countries and other European countries).

This study focuses specifically on primary-school children in an EFL learning context. We chose this context because the acquisition of a foreign language is mandatory in most European countries at a primary level. Furthermore, foreign language learning is one of the priority areas for European cooperation in education, along with transversal key competences and lifelong learning strategies, such as self-regulation strategies, which allow individuals to be better

prepared for contemporary labour markets (European Commission, 2009). In Portugal, learning EFL is optional, not mandatory, which could compromise students' performance in EFL classes. Hence, decided to invest in this curriculum area in order to meet the challenges posed by modern learning and working environments.

What's more, time is a variable, which must be considered when students are expected to acquire a foreign language because the capacity to learn it is reduced as children become older (Dixon, Zhao, Shin, et al., 2012). Thus, it becomes increasingly difficult to learn a foreign language as a native speaker when children reach their teen years (Johnson & Newport, 1989; Mayberry & Locke, 2003). This is also one of the reasons why we proposed to study this age group (8 to 10 years of age). Furthermore, we chose to work with fourth-grade children because it is a transitional grade in Portugal, where children leave primary school and head towards a different system of education where EFL becomes mandatory and the demands of the discipline increase.

Instruments

Interview protocol. This instrument includes questions that ask students about MK, such as how they view themselves as learners of EFL (what their role was); what they think about their class; what they think about how they do in class; and in which ways they learn, independently of liking a task or not. In terms of metacognitive skills, students are asked about how they prepare for their tasks; how they search for and organized information; how they correct their work as they do it; how they evaluated their work; and how they feel they learned (in this particular case on an EFL class, in terms of listening, speaking, reading and writing).

CATOM. The implementation approach (or protocol) of this on-line instrument was based on the principals of dynamic assessment and the mediated learning approach (Ahmed & Pollitt, 2010; Tzuriel & Shamir, 2002). It includes 19 items on a 5 point scale from never (1) to always (5). Higher scores reflect students that reported to have a higher level of MA (including items that tap on MK and MS). This instrument included cartoon images of children studying as a means of motivating the students to respond, but not so many as to distract students or influence their responses. The instrument was constructed to be responded with the guidance of a teacher.

English Task. This task was based on the national EFL curriculum content in Portugal and was developed according to 2 EFL teachers' guidelines. The task included 5 different multiple choice items where students had to identify grammar and vocabulary mistakes and choose the correct response.

Procedures

Development of the CATOM

In order to construct the items for the CATOM, we initially interviewed 20 fourth-grade students in an EFL class and asked them questions regarding their MA with an interview protocol. We obtained responses such as "I'm responsible for the work I do"; "I work well when I study a lot"; and "I follow my teacher's instructions before I start a task". We then tested the Facial Validity and Content Validity of the scale with the participation of 3 fourth-grade students (with a digital audio recorder). This procedure included authorized individual think aloud sessions that integrated spontaneous commentaries and suggestions on the students' behalf, as well as simultaneous cognitive interpellation from the researcher conducting these sessions as each student viewed and responded to the questionnaire (i.e. of question and answer: "Put the question in your own words."; "I know if I'm doing a test correctly or not because of how much I studied before."). Subsequently, we had a focus-group reflection about the scale including all of these three students (i.e. of question and answer: "What is the questionnaire for?"; "This questionnaire is for our teachers to know what we think we do in class, before, during and after tasks"). Two primary school teachers responded to an open-ended question questionnaire about the scale (i.e. of question and answer: "What does the questionnaire measure?"; "The questionnaire allows students to think about their own work in class."). The individual interview guide, the focus-group interview guide, as well as the teacher's open-ended questionnaire were designed according to other studies (Bourque & Fielder, 1995; Dillman, 2000; Fink, 1995).

After making the necessary alterations according to the students' and teachers' comments and suggestions, we had a total of 19 items of MA that tapped on specific issues relating to MK and MS. Essentially, the items we considered as MK included, item 1: "I'm responsible for the work I do" (autonomy belief); item 2: "I am responsible for finishing tasks" (theory about goal); item 4: "I work well when I study a lot" (belief of cognitive functioning); item 6: "I can do a good job" (belief about self); item 10: "I know my ways of learning" (theory about cognitive functioning); item 15: "I work well when the task is easy" (self and task belief); and item 16: "I feel I've learned if I get a good grade" (theory about goal). As for the MS items, we considered item 3: "I follow my teacher's instructions before I start a task" (orientation strategy); item 5: "I make an effort even if I don't like a task" (motivational regulation); item 7: "I make an effort to concentrate" (regulation of cognitive processing); item 8: "I like preparing my work" (motivational regulation and planning strategy); item 9: "I do something I like if I get a good grade" (motivational regulation); item 11: "I like tasks when I am doing them in class" (motivational regulation and monitoring strategy); item 12: "I am interested in tasks because I should pay attention" (motivational regulation); "item 13: I

think about the work I've done" (evaluation strategy); item 14: "I make an effort if I really like the task" (motivational regulation); item 17: "I think about the work I'm going to do before I start" (orientation strategy); item 18: "I think about how I'm going to do my work before I start" (planning strategy); and item 19: "I tell myself I must be interested in assignments" (motivational regulation).

Preliminary Testing of the CATOM

The 19 item CATOM was then delivered on-line in EFL class (with parent, school and student authorization) and was done individually in class by each of the 805 students with teacher guidance. This procedure was followed by all participants. It took students approximately 30 minutes to complete. Students were asked to give an example of the situations or similar situations that had happened to them in order to clarify whether or not they understood the items. If students still had doubts regarding a specific item, they asked either their teacher or the researcher to clarify. The researchers of this study observed the implementation of the instrument, so as to register any important occurrences during each session and to help the students with any doubt that might emerge. Once we gathered the data, we proceeded with an EFA with FACTOR 9.20 (Lorenzo-Seva & Ferrando, 2013) in order to understand the instrument's structure in terms of the number of factors it would yield. Specifically, we were interested in seeing whether separate components would hold for MK and MS or a single unidimensional instrument of MA including both MK and MS. Essentially, if children distinguished between their metacognitive knowledge and skills or if they considered both as one construct of awareness of their metacognition.

Item Response Theory Approach

When we reached an interpretable structure for the instrument, which is described in detail in the results section, we proceeded to apply it a second time to 201 students. These students also performed an English task so as to allow us to assess their performance in EFL in comparison with their performance in the CATOM. As seen in previous studies (Ferreira, Almeida, & Prieto, 2011; 2012), we decided to use a type of statistical analysis that is distinct from the Classical Test Theory for this second analysis, because it would allow us to better understand students' ratings. Specifically, we proceeded with the Rasch analysis with the Winsteps program (Linacre, 2013) in order to assess the unidimensionality of the instrument, as well as to understand how the children had rated their MA. This software allowed us to estimate the students' score on a one-dimensional logit scale and evaluate the properties of the CATOM. Rasch polytomous methodology was adopted to analyze the instrument and the children's ratings. That is, we used the Partial Credit Model (PCM), which is an extension of the Rasch model for polytomous items (Rasch, 1980). Essentially, the PCM for

linear measures of observations of ordinal scales is $\log(P_{nik}/P_{ni(k-1)})/\Theta_n - \beta_i t_{ki}$, where P_{nik} is the probability that person n when encountering item i responds in category k . Accordingly, $P_{ni(k-1)}$ is the probability that the response is in category $k-1$, Θ_n is the ability of person n , β_i is the difficulty (or as proposed in this study, the level of rating) of item i , and t_{ki} is the step calibration in the rating scale threshold (which is defined as the position equivalent to the equal probability of responses in adjacent categories $k-1$ and k (Wright & Masters, 1982). In this study for instance, categories alter from 1 to 5 for MA. The higher score (5) represents overrating (always), whereas the lower score (1) represents underrating (never).

All items were assessed to understand whether they fit the model ($p < .01$) or whether there were items with excessive infit and outfit mean square residuals. That is, we considered removing infit standardized mean squares higher than 1.4 and outfit standardized mean-squares higher than 2.0, as suggested in the literature (Bond & Fox, 2007).

Results

Exploratory Evidence of the CATOM

In a first attempt to interpret the internal structure of the instrument we developed a set of EFA with the data gathered from the 805 participants. Table 1 shows the correlations among all variables and the descriptive statistics. Item scores were uniformly positive correlations (most $r > .30$). Most of the variables were approximately normally distributed, with skewness values less than 2 and kurtosis values less than 5 (Bollen & Long, 1993). Nonetheless, items 14 ($S = -2.080$) and 16 ($S = -2.251$) were negatively skewed. Consistently with Bollen and Long (1993), there is multivariate normality if Mardia's coefficient is lower than $P(P + 2)$, where P is the number of observed variables. In this study, 19 observed variables were used with a Mardia's coefficient for skewness of $65.39 < 19(19 + 2) = 399$ and for kurtosis of $604.09 > 19(19+2) = 399$. Hence, because of our kurtosis values, we used Unweighted Least Squares (ULS) as the method for factor extraction, an estimation method that does not depend on distributional assumptions (Joreskog, 1977). We also used polychoric correlations which are advised when univariate distributions of ordinal items are asymmetric for polytomous items (Brown, 2006; Muthén & Kaplan, 1985; 1992). Furthermore, the data was subjected to the Kaiser-Meyer-Olkin and Bartlett Sphericity test to check for an underlying structure of the data. Essentially, the Kaiser-Meyer-Olkin measure of sampling adequacy was .94, whereas the Bartlett Sphericity was $\chi^2_{171} = 3798.7$ ($p < .001$), demonstrating that the variables were suitable for factor analyses.

Table 1. Item descriptive statistics, slope and threshold parameters and polychoric correlations.

Variable	Mean (SD)	Factor Loadings	Slope and threshold parameters					Polychoric Correlations																	
			a	b ₁	b ₂	b ₃	b ₄	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	4.46(.875)	0.607	0.76	-3.61	-3.05	-1.51	-0.69																		
2	4.25(.988)	0.603	0.75	-3.18	-2.61	-1.27	-0.15	0.40																	
3	4.46(.862)	0.596	0.74	-3.83	-3.19	-1.69	-0.71	0.39	0.32																
4	4.48(.869)	0.613	0.77	-3.47	-2.99	-1.60	-0.76	0.39	0.37	0.31															
5	4.33(.947)	0.623	0.79	-3.29	-2.60	-1.37	-0.35	0.31	0.38	0.41	0.33														
6	4.48(.824)	0.668	0.89	-3.46	-2.95	-1.47	-0.62	0.47	0.40	0.44	0.48	0.36													
7	4.40(.869)	0.704	0.99	-3.17	-2.71	-1.25	-0.40	0.41	0.43	0.42	0.44	0.48	0.42												
8	4.45(.884)	0.620	0.79	-3.54	-2.84	-1.52	-0.64	0.37	0.38	0.37	0.33	0.43	0.34	0.43											
9	4.27(.982)	0.509	0.59	-3.85	-3.31	-1.54	-0.31	0.32	0.27	0.26	0.36	0.30	0.38	0.36	0.30										
10	4.33(.956)	0.638	0.82	-3.17	-2.63	-1.24	-0.37	0.40	0.37	0.36	0.42	0.37	0.42	0.38	0.40	0.30									
11	4.50(.884)	0.615	0.78	-3.46	-2.84	-1.60	-0.85	0.36	0.38	0.39	0.30	0.44	0.41	0.45	0.45	0.21	0.33								
12	4.41(.854)	0.753	1.14	-3.01	-2.56	-1.20	-0.38	0.42	0.45	0.43	0.50	0.49	0.50	0.50	0.43	0.40	0.46	0.48							
13	4.25(.948)	0.628	0.80	-3.39	-2.65	-1.15	-0.13	0.36	0.32	0.37	0.32	0.41	0.36	0.44	0.44	0.33	0.48	0.37	0.47						
14	4.61(.763)	0.738	1.09	-3.26	-2.56	-1.60	-0.89	0.46	0.47	0.44	0.49	0.48	0.47	0.59	0.41	0.38	0.50	0.43	0.57	0.45					
15	4.05(.879)	0.598	0.74	-3.86	-3.22	-0.87	0.52	0.37	0.37	0.31	0.38	0.32	0.45	0.42	0.38	0.29	0.37	0.36	0.45	0.34	0.37				
16	4.59(.834)	0.582	0.71	-3.56	-3.18	-1.90	-1.18	0.37	0.34	0.40	0.37	0.30	0.45	0.31	0.35	0.26	0.37	0.38	0.38	0.33	0.46	0.40			
17	4.05(1.08)	0.567	0.68	-3.01	-2.52	-0.90	0.15	0.33	0.37	0.34	0.32	0.31	0.34	0.42	0.40	0.33	0.36	0.34	0.42	0.39	0.38	0.34	0.31		
18	4.18(.995)	0.578	0.70	-3.38	-2.76	-1.11	-0.00	0.30	0.30	0.31	0.30	0.32	0.39	0.47	0.37	0.29	0.36	0.35	0.46	0.42	0.37	0.35	0.33	0.29	
19	4.14(1.00)	0.643	0.83	-3.11	-2.38	-0.92	0.03	0.36	0.40	0.36	0.38	0.42	0.37	0.41	0.36	0.36	0.45	0.39	0.48	0.41	0.42	0.41	0.40	0.32	0.46

So as to determine the suitable number of factors to retain, various factor retention criteria were applied, specifically, Velicer's MAP test and Horn Parallel analyses. These tests are superior to other standard factor criteria, such as Cattell's Scree test or the Kaiser criterion (O'Connor, 2000). Consistent with the different retention criteria, one factor was obtained (MA) with 42.5% of explained variance. Also, the values of goodness-of-fit ($GFI = .99$), residuals statistics ($RMSR = .037$) and the Guttman-Cronbach's alpha coefficient ($\alpha = .93$) were good in accordance with the literature (McDonald, 1999; Nunnally, 1978; Velicer, 1976). Table 1 also shows the item loadings, as well as the Normal-Ogive Graded Response Model (GRM) parameters, where most items revealed moderate item discrimination. Only item 9 revealed low item discrimination, having scored .591, as indicated in the literature (Baker, 2001). Item discrimination reveals how well an item separates respondents with abilities below the item location from those with abilities above the item location. Hence, we performed the analysis again without item 9 to see how the model would behave (see table 2). Moreover, the item difficulty appears for each item, but not for each category because the category distance is equal due

to the GRM rating scale having the same response options across items. Lastly, because the participants' person-fit indices did not surpass 2.0 (Bond & Fox, 2007), we looked at person reliability and removed 53 participants whose person reliability was $< .70$ and conducted the analysis again. We wanted to check for person reliability because we have polytomous data, rather than binary data, and wanted to avoid any effects of guessing due to the multiple choice format of the questions in the instrument (see Ferrando, 2010). Table 2 shows a comparison between 4 proposed EFA models: (1) with all participants and item 9; (2) with all participants but without item 9; (3) without participants with low individual reliability and with item 9; and (4) without participants with low individual reliability and without item 9. The removal of the participants altered the parameters, although the values presented were still good. The removal of item 9 improved the model essentially in terms of % explained variance (from 42.5% to 43%). In the next section, we present the results of a more detailed analysis using the IRT approach with a different sample which allowed us to confirm the permanence or removal of item 9 and, allowed us to draw more detailed conclusions about the participants' responses.

Table 2. Proposed Unidimensional EFA model parameters of the CATOM

Proposed EFA Model*	Mardia's Coefficient		Kaiser-Meyer-Olkin	Bartlett Sphericity	% Explained Variance	GFI	RMSR	α	Eigenvalues
	S	K							
1	65.39 < 19(19 + 2) = 399	604.09 > 19(19+2) = 399	.94	$\chi^2_{171} = 3798.7 (p < .001)$	42.5%	.99	.037	.93	7.492
2	60.00 < 19(19 + 2) = 399	547.27 > 19(19+2) = 399	.95	$\chi^2_{153} = 3629.6 (p < .001)$	43%	.99	.037	.93	7.233
3	58.48 < 19(19 + 2) = 399	568.21 > 19(19+2) = 399	.94	$\chi^2_{171} = 3153.4 (p < .001)$	39%	.99	.039	.92	6.763
4	53.60 < 19(19 + 2) = 399	514.83 > 19(19+2) = 399	.94	$\chi^2_{153} = 3019.9 (p < .001)$	40%	.99	.038	.92	6.552

* Velicer's Minimum Partial Test used. Horn Parallel Analyses presented same values.

Measuring Perceived MA with the Item Response Theory Approach

We measured the reports of 201 students' MA (CATOM) and their performance in the English task with the Item Response Theory Approach in order to test the unidimensional structure of the instrument and in order to understand whether participants overrated their MA. None of the items showed infit/outfit higher than 1.5, (except for item 9, which had 1.7), as well as χ^2 statistic > 2.00 . We then removed item 9 and carried out the analysis again (see table 3). All items were within the recommended parameters. Item 13 was the easiest or the least reported item with a reported/difficulty level of $-.51$ log, whereas the most difficult or most reported was Item 14 with a reported/difficulty level of $.51$ log. The distribution revealed a narrow range of difficulty ($-.51 < Di < .51$).

We also considered other reliability indicators from the Rasch measures for MA including, Cronbach's alpha, Person Separation Reliability and the Item Separation Reliability. The Person Separation Reliability indicates the proportion of the sample variance which is not explained by the measure

error, while the Item Separation Reliability shows the percentage of item variance that is not explained by the measurement error (Smith, 2001). In this sense, MA revealed a Cronbach's α of .95, a Person Separation Reliability (PSR) of .87, and an Item Separation Reliability (ISR) of .87. These scores indicate good internal consistency reliability (Fox & Jones, 1998) and are higher than the model with item 9 (PSR = .86; ISR = .86). The Person Separation Reliability for MA also reveals, along with the difficulty indicators, that these children may have overrated their awareness of metacognition.

Table 3. IRT parameters of the CATOM

	A	Item Separation Reliability	Person Separation Reliability
19-item Model	.95	.86	.86
18-item Model	.95	.87	.87

Figure 1 is a good visualization of how the children rated their MA. Our results revealed that these children's perceived MA ($\theta = 1.95$) is considerably higher than their achievement in the English task ($\theta = -.89$) represented in figure 2. Additionally, as seen in figure 3, Item 5 appears to

have a considerable level of difficulty when compared with the other items of the English task. This may explain why the mean of the items' level of difficulty is higher than the mean of the children's performance.

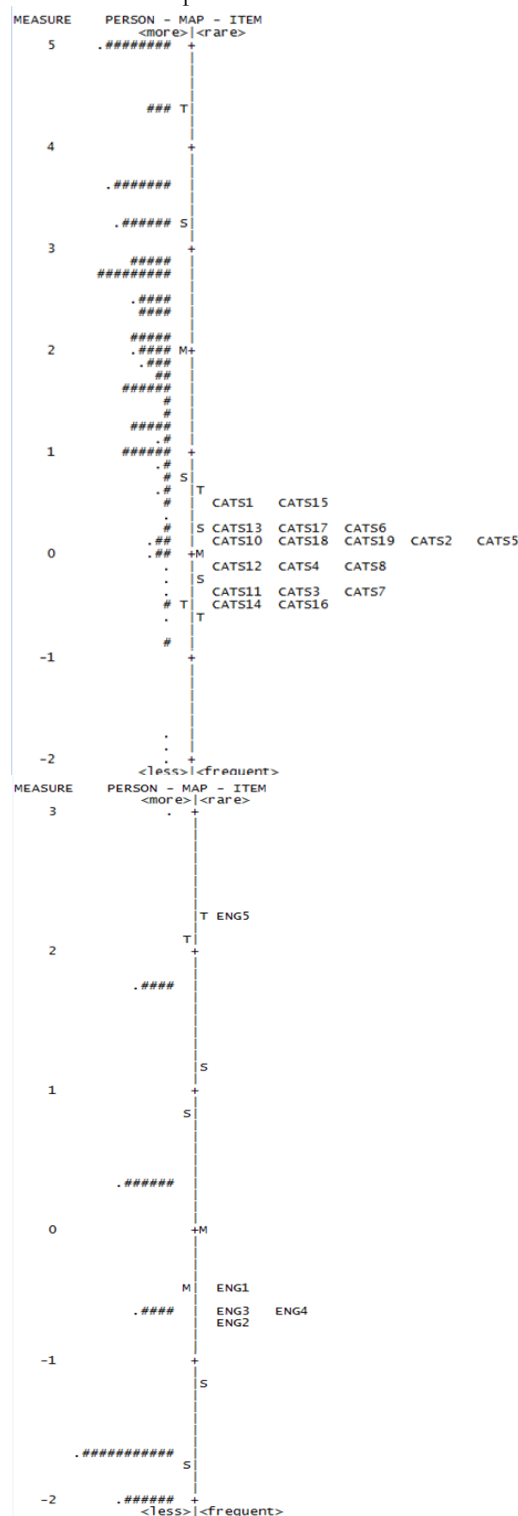


Figure 1. Person-item map for MA Fig 2 Person-item map for the English task.

Discussion

This study presented a preliminary study that proposed an approach of how the person level (MA: MK and MS) of the MASRL model could be measured in fourth-grade children. Because the literature has indicated that it is still unclear and more empirical evidence is needed on how children acquire MA (Efklides, 2011), this study aimed to understand the accuracy with which children reported their metacognitive functioning. Hence, besides testing the initial structure of the instrument with EFA analysis, we chose to use Item Response Theory, which allowed us to calibrate our participants and items on a common scale (De-Mars, 2010; Embretson, 1996). This type of measurement provided an analysis of the interactions between our participants and items, which aided us in interpreting the variables we wanted to measure. Moreover, the interpretations of the items in which participants had a higher probability of dominating was more convenient for our study than group-related ratings. In order to measure children's MA, and because the literature (Wigfield, Klauda, & Cambria, 2011) has indicated a lack of studies and instruments with lower grade levels regarding metacognitive and motivational aspects of self-regulation, we developed the CATOM.

We constructed the CATOM in order to measure how children report their MA of their actions in class. We conclude that it serves its purpose of providing us with information regarding children's metacognitive functioning. We did not expect this instrument to be an event measure and to assess self-regulated learning as a process. Instead, we expected it to be a didactic tool that could give students and teachers information regarding MA and strategy use, as long as students were mediated through its completion. Its expected role, we believe, was confirmed by our results. This instrument is not a process measure to be implemented as students perform actions, but indeed a predictive measure that predicts students' perceptions of their knowledge/tendencies to learn and is to be implemented prior to and subsequently to learning actions.

The psychometric data of the present study can be considered as a preliminary study of the CATOM with a representative sample of 4th grade students. The fact that our results yielded a unidimensional tool indicates that children seem to interpret MK and MS as one only construct (MA), rather than two separate constructs. Also, our results allow us to present the CATOM which may be used to diagnose how children view themselves as metacognitively active agents in their learning process. This instrument could also serve to test hypotheses related to interventions and their expected outcomes in regards to metacognitive and motivational functioning. In fact, this scale could be useful to evaluate the results of an intervention program of Self-regulated Learning in Primary Education, along with other measures, such as diaries, to measure any changes occurring in terms of

students' MA of the learning strategies they use inside classrooms.

In terms of the hypothesis of this study, we feel that the IRT analysis allowed us to interpret the results considering both person and item aspects accurately (De-Mars, 2010; Embretson, 1996). In this sense, results revealed from the reliability results that the item scores were good, including that alpha value ($\alpha = .87$) of these items. From these results we feel that this instrument has potential for future use and testing in other contexts where MA is to be assessed. Our person reliability scores were also good ($\alpha = .87$).

The item difficulty distribution of the CATOM was low in comparison with the students' responses, indicating that the children overrated their MA. In other words, although some studies indicate that children in primary school have the capacity of being consciously aware of themselves and of their thinking processes (Bronson, 2000), our results show that there is a tendency for them to overrate their MA. Results also show that the distribution of the children's responses in the CATOM was higher than the item difficulty distribution. The reverse occurred in the English test, where the item difficulty distribution was higher than the distribution of the children's responses. This leads us to conclude that although children may gain awareness and learn to use strategies intentionally in a self-regulated way, as some studies have stated (Pihlainen-Bednarik & Keinonen, 2010), they still have difficulty in reporting their MA, as there is a tendency of overrating (Allwood, Ask, & Granhag, 2005; Allwood, Innes-Ker, & Fredin, 2008; Shin, Bjorklund, & Beck, 2007). These findings are similar to those of Lipko-Speed (2013), who found that although young children' overconfidence lowers a bit when reporting about past performance with repeated trials of a same task, their reports continue to be inaccurate. Although the author worked with smaller children, she suggests that this may have implications for the future in terms of performance. That is, children can continue to be overconfident and overrate their performance (and in our study specifically, MA) even when confronted with feedback, which can lead them to lower performance levels (Stipek & MacIver, 1989). Hence, the author suggests repeated training to help children learn from past situations and transfer this knowledge to future learning tasks. We agree and recommend that future studies focus on SRL training, where students could focus specifically on their MA with learning diaries for example, as seen in other studies (Schmitz & Perels, 2011).

We believe that in further testing of the CATOM, although we feel that a dynamic assessment approach to its application should continue to be considered, as recommended in other studies (Ahmed & Pollitt, 2010; Tzuriel & Shamir, 2002), a different on-line format of presentation may be applied, including hypothetical situations as examples and images that illustrate these examples. We also feel that this

measure needs further testing in terms of the discriminative validity, by focusing on other contexts (e.g. different country, different schools, different age group, etc...).

In terms of implications for practitioners in the field of education, we tried to contribute to research in the field of metacognition and self-regulated learning by developing and testing an initial factorial validation of a scale that would help psychology researchers and practitioners identify how learners view and report their use of MK and MS, which in turn, could guide teachers in adapting teaching strategies in order to attend to their students' needs. Considering the literature suggests the need for a stronger link between theory and practice regarding the importance of attending to beliefs concerning knowledge and knowing, as well as their influence on strategy use, comprehension, conceptual change, and cognitive processing (Hofer, 2005), we consider that the use of the CATOM by psychology researchers and practitioners can help teachers come a step closer to understanding how students' reports of regulative functioning are important to academic functioning. As Boekaerts (2002) suggests, teachers should have a good awareness of the potential positive and negative beliefs about different topics their students bring into the classroom. This type of knowledge will allow teachers to plan learning activities in coherence with students' belief. Since MK is founded on self-awareness, reflection and monitoring of cognition while the learner is not engaged in learning tasks, we feel that it would be beneficial for teachers to have knowledge about this aspect of their students' metacognitive functioning.

We think that this instrument may also be used in other studies with this particular age group along with other instruments and methods that are more process oriented (Efklides, 2008), such as semi-structured interviews (Zimmerman & Martinez-Pons, 1986), observations of overt behaviour (Turner 1995), traces of mental events and processes (Winne & Perry, 2000), situational manipulations (Rheinberg, Vollmeyer, & Rollett, 2000) and diary keeping (Randi & Corno, 2000). Lastly, this study contributes to the literature on metacognitive aspects of learning because we chose to develop an instrument to measure fourth-grade students' way of reporting their MA through a dynamic assessment approach, considering most research investigating children's metacognitive aspects of self-regulated learning use other assessment methods, such as interviews (Thronsdon, 2010), observations (Whitebread, Coltman, Pasternak, et al., 2009) and tasks (Krebs & Roebbers, 2010; Borkowski & Turner, 1989). We feel that we have made a small contribution to the literature in filling in this gap.

Acknowledgements.- We would like to thank Aristides Ferreira, Anastasia Efklides, Barry Zimmerman, Gerardo Prieto, Karl Stefens and the reviewers of this journal for their pertinent and supportive insights on the development and validation of the CATOM.

References

- Ahmed, A. & Pollitt, A. (2010). The Support Model for interactive assessment. *Assessment in Education: Principles, Policy & Practice*, 17(2), 133–167.
- Allwood, C. M., Ask, K., & Granhag, P. A. (2005). The cognitive interview: effects on the realism in witnesses' confidence in their free recall. *Psychology, Crime & Law*, 11, 183–198.
- Allwood, C. M., Innes-Ker, A., Holmgren, J., & Fredin, G. (2008). Children's and adults' realism in their event-recall confidence in response to free recall and focused questions. *Psychology, Crime & Law*, 14, 529–547.
- Annevirta, T. & Vauras, M. (2006). Development changes of metacognitive skill in elementary school children. *The Journal of Experimental Education* 74, 197–225.
- Baker, F. (2001). The Basics of Item Response Theory. *ERIC Clearinghouse on Assessment and Evaluation*, University of Maryland, College Park, MD.
- Baker, L. (1994). Fostering metacognitive development. In H. W. Reese (Ed.), *Advances in child development and behavior*. Vol. 25 (pp. 201–239). San Diego: Academic Press.
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology*, 25, 729–735.
- Bares, C. B. (2011). Emerging Metacognitive Processes During Childhood: Implications for Intervention Development with Children. *Children and Adolescent Social Work*, 28(4), 291–299. DOI: 10.1007/s10560-011-0233-1
- Barquero, B., Robinson, E. J., & Thomas, G. V. (2003). Children's ability to attribute different interpretations of ambiguous drawings to a naïve vs. a biased observer. *International Journal of Behavioral Development*, 27, 445–456.
- Boekaerts, M. (2002). *Motivation to Learn*. The International Academy of Education and Bureau of Education. Available at <http://www.ibe.unesco.org>.
- Bollen, K. A., & Long, J. S. (1993). *Testing structural equations models*. Newbury Park, CA: Sage.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model*. Mahwah, NJ: Erlbaum.
- Borkowski, J. G., & Turner, L. (1989). Transsituational characteristics of metacognition. In W. Schneider & F. Weinert (Eds.), *Interaction among aptitudes, strategies, and knowledge in cognitive performance* (pp. 159–176). Boston: Springer-Verlag.
- Bourque, L. & Fielder, E. (1995). *How to Conduct Self-Administered and Mail Surveys: Learning Objectives*. Thousand Oaks, CA: Sage.
- Bronson, M. (2000). *Self-regulation in early childhood: nature and nurture*. New York: The Guilford Press.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Burman, E. (1994). *Deconstructing Developmental Psychology*. London and New York: Routledge.
- Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, 65(3), 245–281. doi:10.3102/00346543065003245
- DeMars, C. (2010). Item response theory. Oxford, UK: Oxford University Press.
- Dillman, D. A. (2000). *Mail and Internet Surveys: The Tailored Design Method*. New York: J. Wiley.
- Dixon, Q., Zhao, J., Shin, J., Wu, S., Su, J., Burgess-Brigham, R., Gezer, M., & Snow, C. (2012). What We Know About Second Language Acquisition. A Synthesis From Four Perspectives. *Review of Educational Research*, 82(1), 5–60. doi:10.3102/0034654311433587
- Dweck, C. S. (1999). *Self-Theories: Their Role in Motivation, Personality, and Development*. Philadelphia, PA: Psychology Press Ltd.
- Efklides, A. (2001). Metacognitive experiences in problem solving: Metacognition, motivation, and self-regulation. In A. Efklides, J. Kuhl, & R. M. Sorrentino (Eds.), *Trends and prospects in motivation research* (pp. 297–323). Dordrecht, The Netherlands: Kluwer.
- Efklides, A. (2006). Metacognitive and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1, 3–14. doi:10.1016/j.edurev.2005.11.001
- Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist*, 13(4), 277–287.
- Efklides, A. (2011). Interactions of Metacognition With Motivation and Affect in Self-Regulated Learning: The MASRL Model. *Educational Psychologist*, 46(1), 6–25. 10.1080/00461520.2011.538645
- Embretson, S. E. (1996). Item response theory models and inferential bias in multiple group comparisons. *Applied Psychological Measurement*, 20, 201–212. DOI: 10.1037/1040-3590.8.4.341
- European Commission (2009). Strategic framework for cooperation on education and training. Retrieved from <http://ec.europa.eu/languages/languages-of-europe/>.
- Ferrando, P. (2010). Assessing short-term individual consistency using IRT-based statistics. *Psicologica*, 31, 319–334.
- Ferreira, A. I., Almeida, L., & Prieto, G. (2011). The role of processes and contents in human memory: An item response theory approach. *Journal of Cognitive Psychology*, 23 (7), 873–885.
- Ferreira, A.I., Almeida, L., & Prieto, G. (2012). Construction of a memory battery for computerized administration, using Item Response Theory. *Psychological Reports*, 111(2), 1–25.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp.231–236). Hillsdale, NJ: Erlbaum
- Flavell, J. H. (2002). Development of children's knowledge about the mental world. In W. W. Hartup & R. K. Silbereisen (Eds.), *Growing points in developmental science: An introduction* (pp.102–122). Hove, UK: Psychology Press.
- Flavell, J. H. (2004). Theory-of-Mind Development: Retrospect and Prospect. *Merrill-Palmer Quarterly*, 50(3), 274–290. doi: 10.1353/mpq.2004.0018
- Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, 1, 324–340.
- Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics*. (4th Edition). London: Sage.
- Fink, A. (1995). *How To Ask Survey Questions*. Thousand Oaks, CA: Sage.
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: Critical review and analysis. *Personnel Psychologist*, 39, 292–314.
- Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45(1), 30–45. doi: 10.1037/0022-0167.45.1.30
- Hacker, D. J. (1998). Definitions and Empirical Foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp. 1–23). Lawrence Erlbaum Associates.
- Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of numbers-of-factors rules with simulated data. *Multivariate Behavioral Research*, 17, 193–219.
- Hofer, B. (2005). The legacy and the challenge: Paul Pintrich's contributions to personal epistemology research. *Educational Psychologist*, 40, 95–105.
- Johnson, J. & Newport, E. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Joreskog, K.G. (1977). Factor analysis by least-squares and maximum-likelihood methods. In K. Enslein, A. Ralston, & H.S. Wilf (Eds.), *Statistical Methods for Digital Computers*, pp.125–153, New York: Wiley.
- Kirsh, C. (2012). Developing children's language learner strategies at primary school. *Education* 3–13, 40, (4), 379–399. doi: 10.1080/03004279.2012.691372
- Krebs, S. & Roebbers, C. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *British Journal of Educational Psychology*, 80, 325–340.
- Kuhn, D. & Dean, D. (2004). Metacognition: A Bridge Between Cognitive Psychology and Educational Practice. *Theory Into Practice*, 43(4), 268–273. doi: 10.1353/tip.2004.0047
- Larkin, S. (2010). *Metacognition in Young Children*. NY: Routledge.

- Lipko-Speed, A. (2013). Can young children be more accurate predictors of their recall performance? *Journal of Experimental Child Psychology*, 114, 357–363.
- Mayberry, R. & Lock, E. (2003). Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis. *Brain and Language*, 87, 369–384. doi:10.1016/S0093-934X(03)00137-8
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: LEA.
- Missailidi, P. (2010). Children's Metacognition and Theory of Mind: Bridging the Gap. In A. Efklides & P. Missailidi (Eds.), *Trends and Prospects in Metacognition Research* (pp. 279-291), New York: Springer. DOI 10.1007/978-1-4419-6546-2_13
- Muthén, B., & Kaplan D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171-189.
- Muthén, B., & Kaplan D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45, 19-30.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers*, 32, 396-402.
- Pihlainen-Bednarik, K., & Keinonen, T. (2010). Sixth graders' understanding of their own learning: A case study in environmental education course *International Journal of Environmental & Science Education*, 6(1), 59-78.
- Pintrich, P. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation*, (pp. 451-502). San Diego: Academic Press.
- Randi, J & Corno, L. (2000). Teacher Innovations in Self-Regulated Learning. In M. Boekaerts, P. Pintrich & M. Zeidner (Eds.), *Handbook of Self-Regulation*, (pp 651-685). New York: Academic Press.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press. (Original work published 1960)
- Rheinberg, F., Vollmeyer, R., & Rollett, W. (2000). Motivation and Action in Self-Regulated Learning. In M. Boekaerts, P. Pintrich & M. Zeidner (Eds.), *Handbook of Self-Regulation*, (pp 503-529). New York: Academic Press.
- Rizzo, P., Steinhausen, H.C., & Drechsler, R. (2010). Self-perception of self-regulatory skills in children with attention-deficit/hyperactivity disorder aged 8-10 years. *Attention Deficit and Hyperactivity Disorders*, 2(4), 171-83.
- Schmitz, P., & Perels, F. (2011). Self-monitoring of self-regulation during math homework behaviour using standardized diaries. *Metacognition and Learning*, 6, 255–273. DOI 10.1007/s11409-011-9076-6
- Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, 1, 291–297.
- Schneider, W. (2008). The Development of Metacognitive Knowledge in Children and Adolescents: Major Trends and Implications for Education. *Mind, Brain, and Education*, 2(3), 114-121.
- Schneider, W. & Lockl, K. (2002). The development of metacognitive knowledge in children and adolescents. In Perfect, T. & Schwartz, B. (Eds.), *Applied metacognition*. Cambridge, UK: Cambridge University Press.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351-371.
- Shin, H., Bjorklund, D. F., & Beck, E. F. (2007). The adaptive nature of children's overestimation in a strategic memory task. *Cognitive Development*, 22, 197–212.
- Smith, E. V. Jr (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measure perspective. *Journal of Applied Measurement*, 2(3), 281-311.
- Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology*, 27, 51-79.
- Stipek, D. J., & MacIver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, 60, 521–538.
- Thomas, G.P. & Au Kin Mee, D. (2005). Changing the learning environment to enhance students' metacognition in Hong Kong primary school classroom. *Learning Environments Research*, 8, 221-243.
- Thronsen, I. (2010). Self-regulated learning of basic arithmetic skills: A longitudinal study. *British Journal of Educational Psychology*, 81(4), 558–578. DOI: 10.1348/2044-8279.002008.
- Turner, J.C. (1995). The influence of classroom contexts on young children's motivation for literacy. *Reading Research Quarterly*, 30, 410-441.
- Tzuriel, D. & Shamir, A. (2002). The effects of mediation in computer assisted dynamic assessment. *Journal of Computer Assisted Learning*, 18, 21-32.
- Veenman, M., & Elshout, J. J. (1999). Changes in relation between cognitive and metacognitive skills during acquisition of expertise. *European Journal of Psychology of Education*, 14, 509-523.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
- Whitebread, D., Coltman, P., Pino Pasternak, D., Sangster, C., Grau, V., Bingham, S., Almqed, Q. & Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children, *Metacognition and Learning*, 4(1), 63-85.
- Wigfield, A., Klauda, S., & Cambria, J. (2011). Influences on the Development of Academic Self-regulatory Processes. In B., Zimmerman & D., Schunk (Eds.), *Handbook of Self-regulated Learning and Performance* (pp. 33 - 48). New York, NY: Taylor & Francis.
- Winne, P. & Perry, N. (2000). Measuring Self-Regulated Learning. In M. Boekaerts, P. Pintrich & M. Zeidner (Eds.), *Handbook of Self-Regulation*, (pp 531-566). New York: Academic Press.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: Mesa Press.
- Zimmerman, B. & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, 23, 614–628.

(Article received: 14-10-2013; revised: 03-06-2014; accepted: 10-06-2014)