Presenting the Meta-Performance Test, a Metacognitive Battery based on Performance

Marcio Alexander Castillo Diaz*
National Autonomous University of Honduras, HONDURAS

Cristiano Mauro Assis Gomes
Universidade Federal de Minas Gerais, BRAZIL

Abstract: The self-report and think-aloud approaches are the two dominant methodologies to measure metacognition. This is problematic, since they generate respondent and confirmation biases, respectively. The Meta-Performance Test is an innovative battery, which evaluates metacognition based on the respondent's performance, mitigating the aforementioned biases. The Meta-Performance Test consists of two tests, the Meta-text, which evaluates metacognition in the domain of reading comprehension, and Meta-number, in the domain of arithmetic expressions solving. The main focus of this article is to present the development of the battery, in terms of its conceptual basis, development strategies and structure. Evidence of its content validity is also presented, through the evaluation of three experts in metacognition, two experts in Spanish language, two experts in mathematics and five students who represent the target population. The results of the judges' evaluations attested to the Meta-Performance Test content validity, and the target population declared that both the battery understanding and taking are adequate. Contributions and future research perspectives of the Meta-Performance Test in the field of metacognition are discussed.

Keywords: *Metacognition, assessment, content validity, performance tests.*

Introduction

New work-related and educational paradigms have required students to develop the metacognitive abilities of awareness, monitoring and regulation of their own thinking and learning, which enable them to deal with the current demands for constant learning required by the knowledge and information society (Cardoso et al., 2019; Gomes, 2007a, 2007b; Gomes & Borges, 2009a; Gomes et al., 2014b; Pereira et al., 2019). Metacognitive abilities are relevant predictors of academic performance, as they involve active processes of interaction of the subject with the objects of knowledge (Abdelrahman, 2020; Cai et al., 2019; Cromley & Kunze, 2020). The subject's active interaction, moreover, is a process that articulates and integrates a series of predictors of student's performance, as is the case of learning approaches (Gomes, 2011a, 2013, 2020; Gomes & Golino, 2012b; Gomes et al., 2011, 2020b, 2021; Rodrigues & Gomes, 2020), students' beliefs about the learning process (Alves et al., 2012; Gomes & Borges, 2008a), self-referential cognitions (Costa et al., 2017), and motivation for learning (Gomes & Gjikuria, 2018).

Together with intelligence (Golino & Gomes, 2014; Gomes, 2005, 2010a, 2010b, 2011b, 2012; Gomes & Borges, 2007, 2008b, 2009b, 2009c; Gomes & Golino, 2012a; Martins et al., 2018; Muniz et al., 2016) and the social-economic variables (Gomes et al., 2020a, 2020c, 2020d; Gomes & Jelihovschi, 2019), metacognition also has a prominent place in the prediction of academic performance (Gomes et al., 2014a; Pazeto et al., 2019, 2020; Pires & Gomes, 2018), because it demands from the subject an intense self-regulatory activity. Research has shown that metacognition is linked to regions of the frontal lobe of the brain, and with learning control and management functions (Dinsmore et al., 2008; Morales et al., 2018; Norman et al., 2019).

Regarding the measurement of metacognition, the literature identifies two general methods to evaluate the construct. These methods are classified based on the temporary relationship between the measure used and the performance of the task: (1) offline methods and (2) online methods. The first includes evaluations that occur before or after the performance of the cognitive task; the second refers to evaluations that occur simultaneously to the performance of the cognitive tasks (Akturk & Sahin, 2011; Ohtani & Hisasaka, 2018). The systematic review of Gascoine et al. (2017) and

---

* **Corresponding author:**

Marcio Alexander Castillo Diaz, National Autonomous University of Honduras, Student Affairs Department/ Graduate Program in Psychometrics and Educational Evaluation, Honduras. ✉ marcio.castillo@unah.edu.hn

the meta-analysis of Ohtani and Hisasaka (2018) identify that self-reporting instruments and think-aloud protocols are the major measures used in offline and online methods, respectively.

The virtually full dominance of self-reporting questionnaires and think-aloud protocols to measure metacognition is detrimental to the area of metacognition, as they produce considerable bias. Although the questionnaires have the advantages of enabling a quick and accessible data collection, they demand that respondents have a good perception about their own internal processes. Nevertheless, the literature has shown that this is not usually the case, showing consistent evidence that many people do not have an accurate self-assessment of their own cognitive processes (Abernethy, 2015). In addition to demanding the respondent's accuracy, self-report questionnaires are related to many other biases or problems, such as acquiescence bias, social desirability etc. (Craig et al., 2020; Wetzel et al., 2016).

In turn, think-aloud procedures also potentiate the generation of various and relevant types of biases. Data collection is done through tasks in which the respondents must achieve some goal and report aloud what they are doing at the same time they do the task to achieve the goal proposed. Both the respondents' performance and their speech are recorded and subsequently evaluated by judges so the metacognitive processes or abilities are measured (Greene et al., 2018; Wolcott & Lobczowski, 2021). Considering the intensive evaluation process, studies using this measure involve small samples (e.g., Van der Stel & Veenman, 2008; Veenman & Van Cleef, 2018). Moreover, Priede and Farrall (2010) point out that the need to speak out loud and the possible interference of the evaluator in the respondent's speech; whether it is asking the person to continue "thinking out loud" or ensuring that the respondent is speaking something that can be submitted to analysis, among other elements, produce bias. In addition to this type of evaluator bias, the think-aloud method tends to generate confirmation bias, as metacognitive processes are identified and measured by judges. Das-Smaal (1990, p. 349) warns that, "... real-world features, objects, and events can be categorized in countless different ways. Moreover, our perception is highly selective and therefore, already readily biased."

The use of tests that evaluate metacognition based on respondent's performance may be a viable alternative to mitigate the aforementioned biases. Evaluation using tests that focus on the respondent's own performance does not require judges to evaluate the respondents' performance, as it occurs in the think-aloud protocols, nor does it require respondents to report on their own abilities, as in self-reporting questionnaires. In addition, the evaluation by using performance tests enables to add predictive power and generality of the results, since tests of this nature allow applications in larger samples, without a series of biases found in the questionnaires, which decreases the measurement error and increases the statistical power of the analyses (Castillo, 2018).

There are few tests that evaluate metacognition based on respondent's performance. According to our knowledge, there are only three tests with this characteristic: The Metacognitive Skills and Knowledge Assessment (MSA; Desoete et al., 2001), the Metacognitive Knowledge Test (MKT; Neuenhaus et al., 2011) and the Metacognitive Monitoring Test (also called Reading Monitoring Test or Read Monitoring Test, MMT; Golino & Gomes, 2011; Gomes & Golino, 2014; Gomes et al., 2014a). Both MSA and MKT target primary school students. The MSA aims to evaluate abilities in the domain of metacognitive knowledge (declarative, procedural and conditional) and abilities in the domain of regulation of cognition: prediction, planning, monitoring and judgment. In turn, MKT seeks to evaluate conditional and relational metacognitive knowledge in the domain of reading and mathematics. We found no studies presenting analyses of the structural validity of such tests, by means of factor analysis of items. On the other hand, the MMT shows evidence of convergent, divergent, structural, predictive, and incremental validity for elementary, high school, and higher education students (Castillo, 2018; Gomes et al., 2014a; Gomes & Golino, 2014). Despite the MMT's validity evidence, there are three limitations in the test: (a) It evaluates a single metacognitive ability, monitoring, and only in the context of reading texts; (b) The test has an acceptable, but low, reliability, with Cronbach's alpha between 0.63 and 0.73; (c) the test score derives from the respondent's performance in identifying certain errors in a text and correctly justifying each identification. The test has a space for the respondents to write their justifications. As the score depends on the correct justification for each error identification, the evaluator needs to carefully read each justification, which makes the score generating process relatively slow; moreover, some evaluators may interpret some justifications as correct while others may interpret the same justifications as wrong, generating measurement bias.

In 2019, M. A. Castillo and C. M. A. Gomes, researchers from the Cognitive Architecture Mapping Laboratory (*Laboratório de Investigação da Arquitetura Cognitiva* – LAICO), created the Meta-Performance Test both in Spanish and Brazilian Portuguese. The goal of the authors of this battery involves performing a set of tests capable of overcoming certain limitations of the MSA, MKT and MMT tests, so that it can become, in the future, a useful tool for the evaluation of metacognition. Among the improvements proposed in the Meta-Performance Test, it is worth noting that:

(1) It generates its scores based on data derived from the respondent's performance, without requiring that respondents justify their answers, as in the MMT test.

(2) The battery aims to assess three specific abilities, namely, planning, monitoring and judgment, using a reading comprehension test and an arithmetic expression solving test. This allows to empirically analyze whether these three abilities occur both in the specific domains of reading comprehension and arithmetic expressions solving and in a broad domain that is independent of domain. For example, the battery allows investigating the empirical presence of the reading comprehension monitoring ability, the arithmetic expressions solving monitoring ability, as well as the

monitoring ability regardless of domain. This condition overcomes the MMT limitation, which evaluates the reading comprehension monitoring, but does not permit to estimate the monitoring regardless of the domain.

(3) Since its target population are undergraduate or graduate students, one of the goals of the Meta-Performance Test is to be a tool for the diagnosis of processes involved in the self-regulation of adult learning. This assessment tool seeks to facilitate the design of feasible and concrete interventions that focus on training these abilities and boost the adult audience's ability to think and learn. The literature provides evidence that metacognitive abilities training has a positive impact on the students' academic performance (e.g., Alias & Sulaiman, 2017; Blummer & Kenton, 2014; Perry et al., 2018).

(4) Since the mistakes made by the respondents can be identified in the multiple-choice options, the Meta-Performance Test becomes a useful tool for a more accurate evaluation of the process that leads the respondent to answer in a wrong way. This is a major advance that is in line with areas of knowledge that are focused on the evaluation of the process, such as music therapy (André et al., 2016, 2017, 2020a, 2020b; Rosário et al., 2019; Sampaio et al., 2015).

Considering the arguments presented, the goal of this article is to present the rationale and development of the Meta-Performance Test in detail, showing its conceptual basis and the strategies that supported its development. The article also presents the first evidences of its content validity.

The Meta-Performance Test's Rationale

Conceptual Basis

The Meta-Performance Test takes the consensus built by the field of metacognition as a reference for the definition of its components. There is a relative consensus that metacognition consists basically of two major domains: (1) metacognitive knowledge and (2) control or regulation of cognition (e.g., Baker et al., 2020; Dent & Koenka, 2016; Ohtani & Hisasaka, 2018). The metacognitive knowledge about cognition (also called knowledge about cognition) refers to the individual's knowledge or beliefs about the variables that interact and can affect the course and results of cognitive tasks (Flavell, 1979). It also refers to the individual's knowledge about himself, about his abilities, limitations, difficulties, weaknesses, potentials, internal processes. In other words, the metacognitive knowledge involves the individual's knowledge of his inner world and everything that affects his performance in the external world (Rhodes, 2019).

In turn, the regulation of cognition is the set of abilities that allow the individual to control or regulate his own cognitive activities (Veenman et al., 2006). Different names have been used in the literature for the domain of regulation of cognition (e.g., metacognitive skills, metacognitive skillfulness, or metacognitive learning strategies), which hinders greater conceptual integration among researchers. However, despite differences in nomenclature, the conceptual definitions of the construct are quite similar, if not identical (e.g., Peña-Ayala, 2015; Veenman et al., 2004).

Within the educational context, the domain of regulation of cognition has proven to be significant in terms of academic prediction, with correlations ranging from 0.31 to 0.54 (Ohtani & Hisasaka, 2018). Veenman et al. (2006) linked planning, monitoring and judgment abilities to the domain of regulation of cognition. These abilities are part of this domain because they reflect the online process or "at the time" that it includes awareness and regulation of one's own cognitive operations that occur before, during or after completion of a task. The Meta-Performance Test aims to evaluate these three specific metacognitive abilities.

Planning

The literature provides several definitions for planning. For instance, a classical definition proposed by Owen (1997) says that planning refers to the ability to organize cognition and behavior in time and space, being necessary in situations where an objective must be achieved through a series of intermediate steps, each of which does not necessarily lead directly and individually to the objective. In turn, a contemporary definition is proposed by Oliveira and Nascimento (2014). The authors describe the construct as the capacity to intentionally define and structure actions and resources, aiming to efficiently achieve an objective. Additionally, they created a self-report instrument that aims to measure the different components of the construct.

Li et al. (2015) point out that, compared to other metacognitive abilities, planning is less empirically studied due to the difficulty of separating it from task performance and other components. In what concerns the evaluation of the construct based on performance, the literature indicates the existence of some tests under the tower tasks paradigm, i.e., the Tower of Hanoi (TOH) and the Tower of London (TOL; Georgiou et al., 2017). However, some limitations are observed in relation to these tests. Both TOH and TOL are used mostly in clinical contexts where the objective is to evaluate the occurrence of cognitive impairments in a given patient (e.g., Rodrigues et al., 2019). Because of this condition, the tests tend to generate the ceiling effect, since people who show normal development tend to get a maximum score, which makes it impossible to properly distinguish those people with a higher level of planning (e.g., Kofsky et al., 2014; Oliveira & Nascimento, 2014). In addition, the tests mentioned do not aim to measure only the

planning ability, but a mixed set of abilities, so the score produced by these tests is a planning measure with other cognitive abilities, such as problem solving, working memory or inhibition (Sullivan et al., 2009).

The Meta-Performance Test sees the planning construct as a very specific cognition self-regulation process, i.e., the ability to formulate and/or select a sequence of steps or strategies to solve a task. In this sense, the definition of planning adopted is restricted, because the perspective that the definition of the problem, among other processes, is part of the planning does not fit in this definition. The way the Meta-Performance Test sees it, it is possible to plan with a certain degree of clarity and accuracy and, at the same time, define the problem of a task in an inadequate manner, considering, therefore, that these processes are related, but different abilities.

Monitoring

Monitoring is defined and described in different ways; for example, Rhodes (2019) defines it as the ability to observe, reflect and experience the progress of cognitive processes. In turn, in Nelson and Narens (1996), monitoring is a regulatory element responsible for the flow of information from the object level (cognition) to the target level (metacognition), thus reflecting a cognitive activity self-regulation process. Rhodes' (2019) and Nelson and Narens' (1996) definitions see monitoring as a general metacognitive ability, similar to the definition of the broad domain of regulation of cognition.

Despite the definitions of monitoring as a general metacognitive domain, the Meta-Performance Test approaches the construct as a specific metacognitive ability from the perspective of the error detection paradigm, i.e., as the process that allows the individual to identify flaws or errors at the time of solving a cognitive task (Markman, 1977). The error detection paradigm was originally used in reading comprehension monitoring studies conducted by Markman (1977, 1979). As indicated by Baker (2016), in this paradigm errors or problems are introduced in the texts and various indices are used to determine whether readers notice the problems and try to solve them. The Meta-Performance Test incorporates a similar error detection paradigm to that used in MMT (Gomes & Golino, 2014; Gomes et al., 2014a) and in the methodology proposed in Pires and Gomes (2018).

Judgement

This ability refers to probabilistic evaluation of one's own performance in a task performed (Mihalca et al., 2017). Schraw (2009) proposes a classification of metacognitive judgment according to the time when the evaluation of a task is made. The author calls prospective judgment or prediction the judgment made before the performance of a task. Some typical measures of this type of judgment include the Ease of Learning (EOL), Feeling of Knowing (FOK) and Judgements of Learning (JOL) (e.g., Jemstedt et al., 2017; Taub et al., 2021). On the other hand, concurrent judgment refers to the evaluation that takes place during the process of solving the test items. Thus, this judgment occurs during the test, i.e., right after the respondent finishes to answer each item, judging whether they got the item right or wrong. Finally, the retrospective judgment concerns the assessment of performance after completion of the test. This type of judgment has a more general character, involving the evaluation of performance in the entire test, rather than item by item, as it happens in the concurrent judgment. In addition to classify the judgment by the time it occurs, the judgment is classified as continuous or dichotomous. The continuous judgment involves the use of a confidence scale for correct, usually from 0 to 100, while the dichotomous judgment uses the value of 1 for the confidence for correct and 0 for confidence for error (Schraw, 2008).

In the Meta-Performance Test, the concurrent judgment paradigm is used to evaluate one's own performance during the performance of the tests, analyzing whether or not the performance was successful after the completion of each item. As the judgment is the individuals' own assessment of their performance, in a judgement test the respondents' performance involves precisely the accuracy of their evaluation. In this sense, the degree of accuracy of their perception of whether they get a certain item right or wrong determines the respondents' performance in the judgment test. Thus, the judgment items of the Meta-Performance Test are also based on performance, as are the planning and monitoring items.

Strategies to Develop the Battery

The Meta-Performance Test consists of two tests, the Meta-text, which evaluates metacognition in the domain of reading comprehension, and Meta-number, in the domain of arithmetic expressions solving. Both tests seek to assess planning, monitoring, and judgment metacognitive abilities of the regulation of cognition domain. As pointed out before, the presence of two tests with different cognitive domains allows to evaluate, in terms of latent variables, the aforementioned metacognitive abilities both at specific domains level and regardless of the domains. The development of the tests took as reference the criteria of the American Educational Research Association et al. (2014) through the Standards for Educational and Psychological Testing for the development, validation, application and interpretation of tests.

General Development Strategies

As the identification of errors is not a common task, whether in everyday or academic life, the instructions of the Meta-text and Meta-number tests indicate the objectives involved in the clearest and most understandable way, always presenting at least one example of an item along with a hypothetical respondent's response to that item. With the purpose of properly distinguishing people with low, medium, and high levels of the target abilities, both tests have sets of items that, in theory, bring different levels of difficulty to the respondent. The items are presented in order of level of difficulty, i.e., easier items come first and more difficult items are presented later. Each test item has three commands and each of these commands aims to measure one of the three target metacognitive abilities of the battery. The first command (A) aims to measure planning, while the second command (B) seeks to measure the respondents' judgment regarding their planning. The third command (C) aims to measure monitoring.

Markman's error detection paradigm (Markman, 1977, 1979) was used to evaluate monitoring. Baker (2016) reports that one of the issues associated with the error detection paradigm involves the difficulty the participants have to identify errors, because they assume that the content presented is correct and free of inconsistencies; furthermore, respondents may assume that if the content has inconsistencies, they will be explained later in the text. In the Meta-Performance Test, the command that aims to evaluate monitoring has been specially designed to avoid some problems verified in tasks that use the error detection paradigm. In addition, both the Meta-text and the Meta-number instructions emphasize to the respondent that the texts or arithmetic expressions may contain errors, and that the task of the Command C involves identifying such errors, if they occur in the item.

Specific Development Strategies

The strategies to measure planning (Command A) and monitoring (Command C) were customized according to the content of the domain. Below, the strategies for developing the tasks structure (i.e., items) and the presentation of these two commands in each test will be described.

Meta-Text. This test has 18 items. Each item consists of three fundamental elements: (1) a statement that describes the objective of a hypothetical author; (2) some phrases available to write a text, according to the objective requested; (3) a text written by the hypothetical author, using the phrases available (see Figure 1). The content of the objective, the phrases and the text of each item were carefully prepared, covering a wide range of topics (e.g., society, art, nature, health, technology, and others). In addition, the texts contain words that are part of the respondents' vocabulary and current knowledge so to avoid that they do not perform well due to lack of vocabulary or prior knowledge.

Figure 1. Sample item structure of the Meta-text.

To measure planning, Meta-text Command A prompts the respondent to make a plan to create a text that allows the author's goal to be achieved. The respondent should write, in a proper space, the numbers of the phrases that, mandatorily, should have been selected by the author in order for him to correctly achieve his goal. Respondents who properly plan their reading comprehension process are expected to select solely all the phrases necessary to allow the author to achieve his goal.

In turn, the monitoring evaluation, in Command C, asks the respondent to identify if there are errors in the text written by the author. Errors can be phrases that should not have been included in the text because they do not contribute to achieving the author's goal, or phrases that were not but should have been chosen because they are directly related to the author's goal. In most items, the text written by the author contains errors. The exact number of items with error will not be shown, so not to expose the test answer key. To consider the answer correct, the respondent must have identified all the errors in each text written by the author. It is expected that those respondents who properly monitor their reading process can find the existing errors and consequently get more monitoring commands right.

In theory, the Meta-text contains a balanced number of easy, medium and difficult items. The strategies used to vary the level of difficulty of the items were designed according to the level of abstraction of the texts, objectives and phrases available. An item was considered more abstract than the other if it required a greater generation of logical conclusions or required a greater identification of implicit elements present and necessary for the correct answer to the item.

Meta-Number. This test has 18 items consisting of two elements: (1) An arithmetic expression consisting of natural numbers (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), arithmetic operators [addition (+), subtraction (-), multiplication (x) and division (/)] and an answer preceded by the equals symbol (=); (2) Steps leading to the result shown (See Figure 2).

SAMPLE ITEM	
Arithmetic expression:	**Steps presented for solution:**
$4 + 9 \times 5 - 3 = 26$	**1)** $5 - 3 = 2$ **2)** $4 + 9 \times 2$ **3)** $9 \times 2 = 20$ **4)** $6 + 20$ **5)** $6 + 20 = 26$

COMMAND A. Following the rules, make a plan listing the steps for the correct solution of the arithmetic expression.
Steps:

COMMAND B. Do you reckon you answered correctly to Command A?

Yes: _____ No: _____

COMMAND C. Assess the steps presented in the item.

C.1) Are the steps presented correct? Yes: _____ No: _____

C.2) If they are wrong, identify the specific steps that create mistakes.

Figure 2. Sample item structure of the Meta-number.

To measure planning, the Command A requires that the respondent develop a plan, numbering and presenting, in a given space, the steps necessary for the correct solution of the arithmetic expression. To develop the plan, the instructions bring a set of rules related to the structure and sequence of steps. The rules require that the respondent produce a single operation at each step, correctly integrate the result of the equation with the rest of the arithmetic expression, as well as respect the order and hierarchy of the operators. The rules were carefully created to ensure that there is only one sequence of steps to solve the arithmetic expression correctly. The respondent must strictly reproduce these steps in order to get the command right; therefore, respondents who properly plan their arithmetic expression solving process are expected to produce more steps correctly and get more planning commands right.

To evaluate monitoring in the Meta-number, the command C prompts the respondent to evaluate whether the steps provided in the item itself are correct or not. If the steps are not correct, the respondent must identify all the specific steps that generated an error. Most of the items prepared contained errors in the steps presented. The errors are

related to breaking the rules. Some of the errors involve: (1) disrespecting the hierarchy of operators; (2) performing some incorrect integration step; (3) not following the right order to do the calculations; (4) performing incorrect calculations; (5) omitting calculations; (6) writing some anomalous number.

The test is very careful when providing the instructions for the tasks, showing examples for the correct conduct of the commands, so to avoid that the respondent fails to correctly answer the items due to the lack of understanding of the demands requested by the commands. The strategies to vary the level of difficulty of the items were thought of according to the number of digits and operators that compose each arithmetic expression. In case of Command A, the difficulty varies depending on the number of steps that lead to the correct answer of the arithmetic expression. In turn, in Command C, the difficulty varies according to the type of error and the number of error-making steps.

Methodology

Participants

The analysis of the Meta-Performance Test content validity involved four groups of participants. The first consists of three experts in the metacognition construct, while the second consists of two experts in Spanish language; the third consists of two experts in mathematics, and the fourth group consists of five undergraduate and graduate students (master's degree), representing the target population of the test.

The group of experts in the construct consisted of two psychologists, one with a master's degree in school psychology and the other with a master's degree in developmental psychology, and a pedagogue specializing in psychopedagogy; the experts ranged from 29 to 39 years of age. The group of experts in Spanish language included a master in language studies, aged 28, and a pedagogue specializing in languages and literature, aged 30. The group formed by experts in mathematics consisted of two male mathematics university professors aged 30 and 41. The participants of the target population were all adults (3 female and 2 male), two were undergraduate students of psychology, two were doing a master degree in psychology and one was doing a master degree in neuroscience; they were aged between 19 and 42.

Instrument

Meta-Performance Test. Developed by Castillo and Gomes with the purpose of measuring planning, monitoring, and judgment metacognitive abilities at higher education level.

The test consists of the Meta-text test and the Meta-number test, respectively belonging to the domains of reading comprehension and arithmetic expressions solving. Each item of the two tests consists of three commands (A, B, and C). The items were designed in order to have a balanced level of difficulty (easy, medium and difficult). In addition, the tests are designed so that each one takes no longer than 60 minutes. Figure 1 and Figure 2 show item samples for the Meta-text and the Meta-number, respectively. Readers interested in knowing the complete test can request it from the corresponding author.

Considering that the test seeks to evaluate the metacognition based on the respondents' performance, the answers to commands A (planning) and C (monitoring) are scored as follows: the respondents get a score of one (1) if they answer correctly to the command or a score of zero (0) if they respond incorrectly. In turn, command B measures the respondents' judgment of their own performance in command A. If they think they got the command A of a particular item right, then their judgment score on that item will be 1; otherwise, their score will be 0. To calculate the judgment accuracy, we used as a parameter the tetrachoric correlation of the raw scores of the respondents in command A in relation to their raw scores in command B. A positive and high correlation indicates high judgment accuracy.

Procedures for Data Collection and Analysis

All participants received information regarding the procedures and objectives of the work beforehand and their participation was subject to the signature of a Free and Informed Consent Form (FICF). Table 1 presents the content validity stages and shows in a schematic way the tasks each of the samples (participants in the study) was requested to perform, as well as the data collection and analysis strategies.

Table 1. Steps for the meta-performance test content validity

Step	Sample	Task	Collection	Analysis
Step 1	Experts in the construct	- Respond the assessment protocol so to: 1. Analyze the appropriateness and clarity of the instructions for both the Meta-text and the Meta-number. 2. Analyze the items' answer key.	- Refer the Meta-text and Meta-number assessment protocol to the experts via email. - Interview with the experts via Skype. - Notetaking of the experts' comments during the interview.	1. The degree of appropriateness and clarity of the instructions informed by the judges. 2. The reasons for disagreeing with the answer key.
Step 2	Experts in the content: Spanish and Mathematics	**Experts in Spanish:** - Respond the Meta-text assessment Protocol so to: 1. Analyze the writing as well as the appropriateness and clarity of the instructions. - Logically analyze the arguments of the texts and the items and assess the test answer key. **Experts in mathematics:** 1. Analyze the appropriateness and clarity of the instructions. 2. Analyze the test answer key.	- Refer the Metatext and Meta-number assessment protocol to the experts via email. - Interview with the experts via Skype. - Notetaking of the experts' comments during the interview.	- Check problems pointed out by the experts.
Step 3	Target population	- Semantic analysis of the tests' instructions (appropriateness and clarity). - Answer the test: Meta-text and Meta-number.	- Refer the tests via email.	- Check understanding of the test. - Check the answer key. - Notetaking the tests' completion time. - Preliminary analysis of the items' difficulty level.

Results

Considering that the content validity analysis involved three distinct stages, the results will be presented in this sequence. Table 2 shows a summary of the general evaluations of the experts in construct and in content. The three experts in the construct found that the Meta-text and Meta-number instructions are adequate and clear, and that the items and commands of these tests allow, in theory, the evaluation of planning, monitoring, and judgment abilities.

The assessment of the two experts in Spanish language was in agreement with the test answer key, except for the answers to two specific items (item 4 and item 7). During the interview with the experts, they explained why they disagreed with the answer key for these items. After analyzing the explanation, both items were reformulated, following the experts' suggestions. In addition to the considerations about the test answer key, the experts pointed out a few grammar mistakes in the instrument, which were reviewed and corrected. In turn, the two experts in mathematics found the test instructions appropriate and clear. In addition, the experts agreed with the answer key, except for the answer to an item (item 14). After the analysis of the arguments, the item was reformulated according to the experts' suggestions.

Table 2. Summary of the evaluations of the experts in construct and in content

Sample	General remarks
Experts in the construct	- Expert 1: "The test is well structured and allows to evaluate metacognitive components. The instructions are clear and very exemplified, a variation per item is shown in terms of difficulty level." - Expert 2: "The instructions of each of the commands are directly articulated with the evaluation of planning, monitoring and judgment abilities. Therefore, the requested tasks allow the activation of metacognitive processes. I agree with the justification for the answer key for each item." - Expert 3: "The items allow a novel assessment of metacognitive skills in higher education. The rationale for responses to each item are duly justified."
Experts in Spanish language	**Meta-text** - Expert 1: "The writing and logical arguments of the majority of items are presented correctly. However, it is necessary to review the test answer key of items 4 and 7. The logic of the arguments presented in the answer key of these items may create difficulties for respondents." - Expert 2: "In general, the items present an adequate writing and the arguments for each answer are properly supported. A thematic diversity is shown in each item, which can contribute to keep the interest of the respondents during the test."
Experts in mathematics	**Meta-number** - Expert 1: "The test presents a novel and interesting strategy to evaluate the metacognitive process of solving arithmetic expressions. The steps and rules indicated are correctly and clearly defined. Nevertheless, the answer key for item 14 should be revised, since the steps presented to evaluate planning (Command A) might be incorrect (steps 3 and 4)." - Expert 2: "The test requires prior knowledge of mathematics commensurate with the level of higher education. This minimum knowledge is consistent with the target group of the test. I agree with the test answer key, except for item 14."

The Meta-text and Meta-number evaluation by the target population was favorable. Semantic analysis indicated that the test instructions are understandable and the test is feasible. Table 3 displays a preliminary analysis of the items' difficulty level. As can be seen, both the Meta-text and the Meta-number contain items considered easy, average and difficult. It was observed that the mistakes made by the participants were specifically caused by difficulties inherent to the challenges of the items themselves. With regard to test completion time, respondents took an average of 46 minutes to complete the Meta-text and 71 minutes to complete the Meta-number. In view of the time required to answer the items of the Meta-number, the authors think it necessary to reduce to 14 the number of items in this test. This is supported by the fact that four of the five members of the target population reported to be answering item number 15 at minute 60 of the Meta-number test.

Table 3. Preliminary analysis of the items' difficulty level

Level of difficulty	Meta-text (items)	Meta-number (items)
Easy	1, 2, 4, 6, 16	1, 2, 3, 4, 5, 6, 7
Average	9, 10, 11, 12, 13, 14, 18	8, 9, 10, 11, 12
Difficult	3, 5, 7, 8, 15, 17	13, 14, 15, 16, 17, 18

Discussion

This study aimed to present the rationale and development strategies of the Meta-Performance Test. In addition, evidence of this test's content validity was presented. As a result of this paper, some implications to the field of metacognition studies can be pointed out.

The first concerns the presentation of the measurement issue in the area of metacognition. Self-report instruments and think-aloud protocols, despite being the most widely used to measure metacognitive abilities, generate considerable

respondent and confirmation biases, respectively. These biases have been presented in previous research (e.g., Abernethy, 2015; Craig et al., 2020; Priede & Farrall, 2010). The use of performance tests is an alternative to deal with the aforementioned biases. However, the lack of this type of measurement is an important limitation of the area of metacognition. Consequently, this article detailed the Meta-Performance Test, since it allows to evaluate metacognitive abilities based on the respondent's performance. The objective of this article is in line with the review of Ohtani and Hisasaka (2018), who emphasize the need to develop innovative measures that assess specific metacognitive abilities, mainly within the domain of the regulation of cognition.

A second implication involves the presentation of initial evidence of the test's content validity, which has gone through the analysis and scrutiny of experts in the construct, experts in the content, and representatives of the target population. In other words, the test shows evidence that reinforces its theoretical assumptions that its items are markers of the measurement of the target abilities and the tests are feasible. According to the American Educational Research Association et al. (2014), the evidence based on test content is part of the various sources which contributes to generate accumulated evidence of validity. In the development of new instruments, this evidence acquires greater importance since it allows to study the relationship between the content of a test and the construct to be measured.

Finally, the structure of Meta-Performance Test involves the presence of two tests, each of them in a different cognitive domain, allowing the measurement of both metacognitive abilities specific to these domains and the measurement of these abilities, but regardless of the domain. The evidence provided by Rouault et al. (2018) indicates that metacognition not only operates as a domain-general resource applied over cognitive tasks, but also differently depending on the nature of these tasks.

Conclusion

Metacognition includes a set of abilities with important applications in the educational and professional contexts demanded by today's society. Developing objective evaluation measures that allow the nature of the construct to be assessed is highly relevant for conducting diagnostic and intervention processes. The almost exclusive predominance of self-report and think-aloud measurements procedures is a limitation that needs to be overcome and the Meta-Performance Test is a promising alternative.

Recommendations

This study presents only initial evidence of the Meta-Performance Test content validity, and new investigations are necessary for this test to present robust evidence of validity and become available for general use. This study is only the initial part of a broad set of necessary validity studies. Future research should aim to search for evidence regarding the Meta-Performance Test structural validity and external validity, allowing to generate more robust evidence on the instrument's construct validity. If this evidence is favorable, then this test may become an instrument of assessment available to be used by professionals who want a metacognition measurement.

Limitations

Despite that the Meta-Performance Test is intended to measure a set of metacognitive abilities within the domain of regulation of cognition (such as planning, monitoring and judgment), there are other abilities that some authors identify and which are not part of the test (e.g., prediction, orientation, adjusting, debugging, etc.). Moreover, considering that university students represent the target population of the test, major adaptations should be made on items content in case to be validated at other educational levels. These adaptations are of relevance mainly in primary and lower secondary education. Finally, this article focuses only on the presentation and content validity of the test, therefore, empirical evidence is needed about the structure of the test on different samples and cultures.

Authorship Contribution Statement:

M.A.C. Diaz: Conceptualization, design, data acquisition, data analysis, interpretation, drafting manuscript, critical revision of manuscript, final approval. C.M.A. Gomes: Conceptualization, design, data analysis, interpretation, drafting manuscript, critical revision of manuscript, supervision, final approval.

References

Abdelrahman, R. M. (2020). Metacognitive awareness and academic motivation and their impact on academic achievement of Ajman University students. *Heliyon*, *6*(9), 1–8. https://doi.org/10.1016/j.heliyon.2020.e04192

Abernethy, M. (2015). Self-reports and observer reports as data generation methods: An assessment of issues of both methods. *Universal Journal of Psychology*, *3*(1), 22–27. https://doi.org/10.13189/ujp.2015.030104

Akturk, A., & Sahin, I. (2011). Literature review on metacognition and its measurement. *Procedia Social and Behavioral Sciences*, *15*, 3731–3736. https://doi.org/10.1016/j.sbspro.2011.04.364

Alias, M., & Sulaiman, N. (2017). Development of metacognition in higher education: Concepts and strategies. In E. Railean, A. Elçi, & A. Elçi (Eds.), *Metacognition and successful learning strategies in higher education. Advances in higher education and professional development (AHEPD) book series* (pp. 22-42). Information Science Reference.

Alves, F. A., Flores, R. P., Gomes, C. M. A., & Golino, H. F. (2012). Preditores do rendimento escolar: inteligência geral e crenças sobre ensino-aprendizagem [Predictors of school performance: general intelligence and beliefs about teaching-learning]. *E-PSI Journal/ Revista E-PSI, 1*, 97-117.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education-USA (2014). *Standards for educational and psychological testing*. AERA Publications.

André, A. M., Gomes, C. M. A., & Loureiro, C. M. V. (2016). Escalas Nordoff Robbins: uma revisão bibliográfica [Nordoff Robbins Scales: a literature review]. *Percepta, 3*(2), 117-131. https://doi.org/10.34018/2318-891X.3(2)117-131

André, A. M., Gomes, C. M. A., & Loureiro, C. M. V. (2017). Equivalência de itens, semântica e operacional da versão brasileira da escala Nordoff Robbins de comunicabilidade musical [Equivalence of items, semantics and operations of the Brazilian version of the Nordoff Robbins musical communication scale]. *OPUS, 23*(2), 197-215. https://doi.org/10.20504/opus2017b2309

André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2020a). Confiabilidade interexaminadores da Escala de relação criança-terapeuta na experiência musical coativa para validação no contexto brasileiro [Inter-rater reliability of the child-therapist relationship scale in coactive musical experience for validation in the Brazilian context]. *Hodie, 20*(1), 1-18. https://doi.org/10.5216/mh.v20.64243

André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2020b). Equivalência de itens, semântica e operacional da "Escala de Musicabilidade: Formas de Atividade, Estágios e Qualidades de Engajamento" [Equivalence of items, semantics and operations of the "Musicality Scale: Forms of Activity, Stages and Qualities of Engagement"]. *Orfeu, 5*(2), 1–22. https://doi.org/10.5965/2525530405022020e0010

Baker, L. (2016). The Development of Metacognitive Knowledge and Control of Comprehension. Contributors and Consequences. In K. Mokhtari (Ed.), *Improving Reading Comprehension through Metacognitive Reading Strategies Instruction* (pp. 01-31). Rowman & Littlefield.

Baker, L., Millman, Z. B., & Singer Trakhman, L. M. (2020). How the construct of metacognition has contributed to translational research in education, mental health, and beyond. *Translational Issues in Psychological Science, 6*(1), 1-7. https://doi.org/10.1037/tps0000225

Blummer, B., & Kenton, J. (2014). *Improving Student Information Search: A metacognitive approach*. Elsevier.

Cai, Y., King, R. B., Law, W., & McInerney, D. M. (2019). Which comes first? Modeling the relationships among future goals, metacognitive strategies and academic achievement using multilevel cross-lagged SEM. *Learning and Individual Differences, 74*, 101750. https://doi.org/10.1016/j.lindif.2019.06.004

Cardoso, C. O., Seabra, A. G., Gomes, C. M. A., & Fonseca, R. P. (2019). Program for the neuropsychological stimulation of cognition in students: impact, effectiveness, and transfer effect on student cognitive performance. *Frontiers in Psychology, 10*, 1-16. https://doi.org/10.3389/fpsyg.2019.01784

Castillo, M. (2018). *Monitoring and intelligence as predictors of general and specific academic performance in higher education* [Master's thesis, Federal University of Minas Gerais, Brasil]. UFMG Digital Archive. https://repositorio.ufmg.br/handle/1843/BUOS-B3PJKN

Costa, B. C. G., Gomes, C. M. A., & Fleith, D. S. (2017). Validade da Escala de cognições acadêmicas autorreferentes: autoconceito, autoeficácia, autoestima e valor [Validity of the Self-Referential academic cognition scale: self-concept, self-efficacy, self-esteem and value]. *Psychological Assessment/ Avaliação Psicológica, 16*(1), 87-97. https://doi.org/10.15689/ap.2017.1601.10

Craig, K., Hale, D., Grainger, C., & Stewart, M. E. (2020). Evaluating metacognitive self-reports: systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning, 15*(2), 155-213. https://doi.org/10.1007/s11409-020-09222-y

Cromley, J. G., & Kunze, A. J. (2020). Metacognition in education: Translational research. *Translational Issues in Psychological Science, 6*(1), 15-20. https://doi.org/10.1037/tps0000218

Das-Smaal, E. A. (1990). Biases in categorization. *Advances in Psychology, 68*, 349–387. https://doi.org/10.1016/S0166-4115(08)61332-1

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review, 28*(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities, 34*(5), 435–447. https://doi.org/10.1177/002221940103400505

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*(4), 391–409. https://doi.org/10.1007/s10648-008-9083-6

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

Gascoine, L., Higgins, S., & Wall, K. (2017). The assessment of metacognition in children aged 4–16 years: A systematic review. *Review of Education, 5*(1), 3–57. https://doi.org/10.1002/rev3.3077

Georgiou, G., Li, J., & Das, J. P. (2017). Tower of London: What level of planning does it measure? *Psychological Studies, 62*(3), 261-267. https://doi.org/10.1007/s12646-017-0416-8

Golino, H. F., & Gomes, C. M. A. (2011). Preliminary internal validity evidences of two Brazilian Metacognitive Tests. *International Journal of Testing, 26*(1), 11-12. *International Test Commission.* https://www.intestcom.org/files/ti26.pdf

Golino, H. F., & Gomes, C.M.A. (2014). Psychology data from the "BAFACALO project: The Brazilian Intelligence Battery based on two state-of-the-art models – Carroll's Model and the CHC model". *Journal of Open Psychology Data, 2*(1), e6. https://doi.org/10.5334/jopd.af

Gomes, C. M. A. (2005). *Uma análise dos fatores cognitivos mensurados pelo Exame Nacional do Ensino Médio (ENEM) [An analysis of the cognitive factors measured by the National High School Exam (ENEM)]* [Doctoral Thesis, Federal University of Minas Gerais]. UFMG Digital Archive. http://hdl.handle.net/1843/FAEC-85RJNN

Gomes, C. M. A. (2007a). *Apostando no desenvolvimento da inteligência; em busca de um novo currículo educacional para o desenvolvimento do pensamento humano* [Betting on the development of intelligence; in search of a new educational curriculum for the development of human thought]. Lamparina.

Gomes, C. M. A. (2007b). Softwares educacionais podem ser instrumentos psicológicos [Educational software can be psychological tools]. *School and Educational Psychology/ Psicologia Escolar e Educacional, 11(*2), 391-401. https://doi.org/10.1590/S1413-85572007000200016

Gomes, C. M. A. (2010a). Avaliando a avaliação escolar: notas escolares e inteligência fluida [Evaluating school evaluation: school grades and fluid intelligence]. *Psychology in Study/ Psicologia em Estudo, 15*(4), 841-849. https://doi.org/10.1590/S1413-73722010000400020

Gomes, C. M. A. (2010b). Estrutura fatorial da Bateria de Fatores Cognitivos de Alta-Ordem (BaFaCalo) [Factorial Structure of the High-Order Cognitive Factors Battery (BaFaCalo)]. *Psychological Assessment/ Avaliação Psicológica, 9*(3), 449-459.

Gomes, C. M. A. (2011a). Abordagem profunda e abordagem superficial à aprendizagem: diferentes perspectivas do rendimento escolar [Deep approach and superficial approach to learning: different perspectives of school performance]. *Psychology: Reflection and Criticism/ Psicologia: Reflexão e Crítica, 24*(3), 438-447. https://doi.org/10.1590/S0102-79722011000300004

Gomes, C. M. A. (2011b). Validade do conjunto de testes da habilidade de memória de curto-prazo (CTMC) [Validity of the short-term memory ability test set]. *Psychology Studies (Natal)/ Estudos de Psicologia (Natal), 16*(3), 235-242. https://doi.org/10.1590/S1413-294X2011000300005

Gomes, C. M. A. (2012). Validade de construto do conjunto de testes de inteligência cristalizada (CTIC) da bateria de fatores cognitivos de alta-ordem (BaFaCAlO) [Construct validity of the set of crystallized intelligence tests (CTIC) of the battery of high-order cognitive factors (BaFaCAlO)]. *General: Interinstitutional Journal of Psychology/ Gerais: Revista Interinstitucional de Psicologia, 5*(2), 294-316.

Gomes, C. M. A. (2013). A construção de uma medida em abordagens de aprendizagem [The construction of a measure of learning approaches]. *Psico (PUCRS-Online), 44*(2), 193-203.

Gomes, C. M. A. (2020). Programa de enriquecimento istrumental: evidências de eficácia para intervenção cognitiva [Instrumental enrichment program: evidence of effectiveness for cognitive intervention]. In M. Mansur-Alves & J. B. Lopes-Silva (Eds.), *Intervenção cognitiva: dos conceitos às práticas baseadas em evidências para diferentes aplicações [Cognitive intervention: from concepts to evidence-based practices for different applications]* (pp. 621-639). T.Ser.

Gomes, C. M. A., Amantes, A., & Jelihovschi, E.G. (2020a). Applying the regression tree method to predict students' science achievement. *Trends in Psychology, 28*(1), 99-117. https://doi.org/10.9788/s43076-019-00002-5

Gomes, C. M. A., Araujo, J., & Jelihovschi, E. G. (2020b). Approaches to learning in the non-academic context: construct validity of Learning Approaches Test in Video Game (LAT-Video Game). *International Journal of Development Research, 10*(11), 41842-41849. https://doi.org/10.37118/ijdr.20350.11.2020

Gomes, C. M. A., & Borges, O. N. (2007). Validação do modelo de inteligência de Carroll em uma amostra brasileira [Validation of the Carroll intelligence model in a Brazilian sample]. *Psychological Assessment/ Avaliação Psicológica, 6*(2), 167-179.

Gomes, C. M. A., & Borges, O. N. (2008a). Avaliação da validade e fidedignidade do instrumento crenças de estudantes sobre ensino-aprendizagem (CrEA) [Evaluation of the validity and reliability of the instrument students' beliefs about teaching-learning]. *Science & Cognition/ Ciências & Cognição, 13*(3), 37-50.

Gomes, C. M. A., & Borges, O. (2008b). Qualidades psicométricas de um conjunto de 45 testes cognitivos [Psychometric qualities of a set of 45 cognitive tests]. *Fractal: Journal of Psychology/ Fractal: Revista de Psicologia, 20*(1), 195-207. https://doi.org/10.1590/S1984-02922008000100019

Gomes, C. M. A. & Borges, O. N. (2009a). O ENEM é uma avaliação educacional construtivista? Um estudo de validade de construto [Is ENEM a constructivist educational assessment? A construct validity study]. *Educational Evaluation Studies/ Estudos em Avaliação Educacional, 20*(42), 73-88. https://doi.org/10.18222/eae204220092060

Gomes, C. M. A., & Borges, O. (2009c). Qualidades psicométricas do conjunto de testes de inteligência fluida [Psychometric qualities of the fluid intelligence test suite]. *Psychological Assessment/ Avaliação Psicológica, 8*(1), 17-32.

Gomes, C. M. A., Fleith, D. S., Marinho-Araujo, C. M., & Rabelo, M. L. (2020c). Predictors of students' mathematics achievement in secondary education. *Psychology: Theory and Research/ Psicologia: Teoria e Pesquisa, 36*(1), e3638. https://doi.org/10.1590/0102.3772e3638

Gomes, C. M. A., & Gjikuria, E. (2018). Structural Validity of the School Aspirations Questionnaire (SAQ). *Psychology: Theory and Research/ Psicologia: Teoria e Pesquisa, 34*(1), 1-11. https://doi.org/10.1590/0102.3772e3438

Gomes, C. M. A., & Golino, H. F. (2012a). O que a inteligência prediz: diferenças individuais ou diferenças no desenvolvimento acadêmico? [What does intelligence predict: individual differences or differences in academic development?]. *Psychology: Theory and Practice/ Psicologia: Teoria e Prática, 14*(1), 126-139.

Gomes, C. M. A., & Golino, H. F. (2012b). Validade incremental da Escala de Abordagens de Aprendizagem (EABAP) [Incremental validity of the Learning Approach Scale (EABAP)]. *Psychology: Reflection and Criticism/ Psicologia: Reflexão e Crítica, 25*(4), 400-410. https://doi.org/10.1590/S0102-79722012000400001

Gomes, C., & Golino, H. (2014). Self-reports on students' learning processes are academic metacognitive knowledge. *Psychology: Reflection and Criticism/ Psicologia: Reflexão e Crítica, 27*(3), 472-480. https://doi.org/10.1590/1678-7153.201427307

Gomes, C., Golino, H., & Menezes, I. (2014a). Predicting school achievement rather than intelligence: Does metacognition matter? *Psychology, 5*(9), 1095–1110. https://doi.org/10.4236/psych.2014.59122

Gomes, C. M. A., Golino, H. F., Pinheiro, C. A. R., Miranda, G. R., & Soares, J. M. T. (2011). Validação da Escala de Abordagens de Aprendizagem (EABAP) em uma amostra Brasileira [Validation of the Learning Approaches Scale (EABAP) in a Brazilian sample]. *Psychology: Reflection and Criticism/ Psicologia: Reflexão e Crítica, 24*(1), 19-27. https://doi.org/10.1590/S0102-79722011000100004

Gomes, C. M. A., Golino, H. F., Santos, M. T., & Ferreira, M. G. (2014b). Formal-Logic Development Program: Effects on Fluid Intelligence and on Inductive Reasoning Stages. *British Journal of Education, Society & Behavioural Science, 4*(9), 1234-1248. https://doi.org/10.9734/BJESBS/2014/10757

Gomes, C. M. A., & Jelihovschi, E. (2019). Presenting the regression tree method and its application in a large-scale educational dataset. *International Journal of Research & Method in Education, 43*(2), 201-221. https://doi.org/10.1080/1743727X.2019.1654992

Gomes, C. M. A., Lemos, G. C., & Jelihovschi, E. G. (2020d). Comparing the predictive power of the CART and CTREE algorithms. *Psychological Assessment/ Avaliação Psicológica, 19*(1), 87-96. https://doi.org/10.15689/ap.2020.1901.17737.10

Gomes, C. M. A., Linhares, I. S., Jelihovschi, E. G., & Rodrigues, M. N. S. (2021). Introducing rationality and content validity of SLAT-Thinking. *International Journal of Development Research, 11*(1), 43264-43272. https://doi.org/10.37118/ijdr.20586.01.2021

Greene, J. A., Deekens, V. M., Copeland, D. Z., & Yu, S. (2018). Capturing and modeling self-regulated learning using think-aloud protocols. In D. H. Schunk & J. A. Greene (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp. 323–337). Routledge/Taylor & Francis Group.

Jemstedt, A., Schwartz, B. L., & Jönsson, F. U. (2017). Ease-of-learning judgments are based on both processing fluency and beliefs. *Memory, 26*(6), 807-815. https://doi.org/10.1080/09658211.2017.1410849

Kofsky, E., Friedman, S., & Wallner-Allen, K. (2014). What do they really measure? A comparative analysis of planning tasks. In Friedman, S., & Kofsky, E. (Eds.), *The Developmental Psychology of Planning: Why, How, and When Do We Plan?* (pp. 127-156). Psychology Press.

Li, J., Zhang, B., Du, H., Zhu, Z., & Li, Y. M. (2015). Metacognitive planning: Development and validation of an online measure. *Psychological Assessment, 27*(1), 260-271. https://doi.org/10.1037/pas0000019

Markman, E. M. (1977). Realizing that You Don't Understand: A Preliminary Investigation. *Child Development, 48*(3), 986-992.

Markman, E. M. (1979). Realizing that You Don't Understand: Elementary School Children's Awareness of Inconsistencies. *Child Development, 50*(3), 643-655. https://doi.org/10.2307/1128929

Martins, A. A., Gomes, C. M. A., Alves, A. F., & Almeida, L. S. (2018). The structure of intelligence in childhood: age and socio-familiar impact on cognitive differentiation. *Psychological Reports, 121*(1), 79-92. https://doi.org/10.1177/0033294117723019

Mihalca, L., Mengelkamp, C., & Schnotz, W. (2017). Accuracy of metacognitive judgments as a moderator of learner control effectiveness in problem-solving tasks. *Metacognition and Learning, 12*(3), 357–379. https://doi.org/10.1007/s11409-017-9173-2

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *The Journal of Neuroscience, 38*(14), 3534–3546. https://doi.org/10.1523/jneurosci.2360-17.2018

Muniz, M., Gomes, C. M. A., & Pasian, S. R. (2016). Factor structure of Raven's Coloured Progressive Matrices. *Psico-USF, 21*(2), 259-272. https://doi.org/10.1590/1413-82712016210204

Nelson, T., & Narens, L. (1996). Why investigate Metacognition? In J. Metcalfe & A. P. Shimamura (Ed.), *Metacognition. Knowing about knowing* (pp. 1-27). MIT Press.

Neuenhaus, N., Artelt, C., Lingel, K., & Schneider, W. (2011). Fifth graders metacognitive knowledge: General or domain-specific? *European Journal of Psychology of Education, 26*(2), 163–178. https://doi.org/10.1007/s10212-010-0040-7

Norman, E., Pfuhl, G., Sæle, R. G., Svartdal, F., Låg, T., & Dahl, T. I. (2019). Metacognition in Psychology. *Review of General Psychology, 23*(4), 403-424. https://doi.org/10.1177/1089268019883821

Ohtani, K., & Hisasaka, T. (2018). Beyond intelligence: A meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacognition and Learning, 13*(2), 179–212. https://doi.org/10.1007/s11409-018-9183-8

Oliveira, A., & Nascimento, E. (2014). Construção de uma escala para avaliação do planejamento cognitivo [Construction of a cognitive planning assessment scale]. *Psychology: Reflection and Criticism/ Psicologia: Reflexão e Crítica, 27*(2), 209-218. https://doi.org/10.1590/1678-7153.201427201

Owen, A. (1997). Cognitive planning in humans: neuropsychological, neuroanatomical and neuropharmacological perspectives. *Progress in Neurobiology, 53*(4), 431-50. https://doi.org/10.1016/s0301-0082(97)00042-7

Peña-Ayala, A. (2015). *Metacognition: Fundaments, Applications, and Trends. A Profile of the Current State-Of-The-Art.* Springer.

Pazeto, T. C. B., Dias, N. M., Gomes, C. M. A., & Seabra, A. G. (2019). Prediction of arithmetic competence: role of cognitive abilities, socioeconomic variables and the perception of the teacher in early childhood education. *Psychology Studies/ Estudos de Psicologia, 24*(3), 225-236. https://doi.org/10.22491/1678-4669.20190024

Pazeto, T. C. B., Dias, N. M., Gomes, C. M. A., & Seabra, A. G. (2020). Prediction of reading and writing in elementary education through early childhood education. *Psychology: Science and Profession/ Psicologia: Ciência e Profissão, 40*, 1-14. https://doi.org/10.1590/1982-3703003205497

Pereira, B. L. S., Golino, M. T. S., & Gomes, C. M. A. (2019). Investigando os efeitos do Programa de Enriquecimento Instrumental Básico em um estudo de caso único [Investigating the effects of the Basic Instrumental Enrichment Program in a single case study]. *European Journal of Education Studies, 6*(7), 35-52. https://doi.org/10.5281/zenodo.3477577

Perry, J., Lundie, D., & Golder, G. (2018). Metacognition in schools: What does the literature suggest about the effectiveness of teaching metacognition in schools? *Educational Review, 71*(4), 483-500. https://doi.org/10.1080/00131911.2018.1441127

Pires, A. A. M., & Gomes, C. M. A. (2018). Proposing a method to create metacognitive school exams. *European Journal of Education Studies, 5*(8), 119-142. https://doi.org/10.5281/zenodo.2313538

Priede, C., & Farrall, S. (2010). Comparing results from different styles of cognitive interviewing: 'Verbal probing' vs. 'thinking aloud'. *International Journal of Social Research Methodology, 14*(4), 271–287. https://doi.org/10.1080/1:3645579.2010.523187

Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology, 46*(2), 168-175. https://doi.org/10.1177/0098628319834381

Rodrigues, C. L., Rocca, C. C. A., Serafim, A., Santos, B., & Asbahr, F. R. (2019). Impairment in planning tasks of children and adolescents with anxiety disorders. *Psychiatry Research, 274*, 243-246. https://doi.org/10.1016/j.psychres.2019.02.049

Rodrigues, M. N. S., & Gomes, C. M. A. (2020). Testing the hypothesis that the deep approach generates better academic performance. *International Journal of Development Research, 10*(12), 42925-42935. https://doi.org/10.37118/ijdr.20579.12.2020

Rosário, V. M., Gomes, C. M. A., & Loureiro, C. M. V. (2019). Systematic review of attention testing in allegedly "untestable" populations. *International Journal of Psychological Research and Reviews, 2*(19), 1-21. https://doi.org/10.28933/ijprr-2019-07-1905

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging. *Personality Neuroscience, 1*, 1-13. https://doi.org/10.1017/pen.2018.16

Sampaio, R. T., Loureiro, C. M. V., & Gomes, C. M. A. (2015). A Musicoterapia e o transtorno do espectro do autismo: uma abordagem informada pelas neurociências para a prática clínica [Music therapy and autism spectrum disorder: an informed approach by neurosciences for clinical practice]. *Scholarly Music Journal/ Per Musi, 32*, 137-170. https://doi.org/10.1590/permusi2015b3205

Schraw, G. (2008). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*(1), 33-45. https://doi.org/10.1007/s11409-008-9031-3

Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Handbook of metacognition in education* (pp. 415-429). Routledge/ Taylor & Francis Group.

Sullivan, J. R., Riccio, C. A., & Castillo, C. L. (2009). Concurrent validity of the tower tasks as measures of executive function in adults: a meta-analysis. *Applied Neuropsychology, 16*(1), 62–75. https://doi.org/10.1080/09084280802644243

Taub, M., Azevedo, R., Rajendran, R., Cloude, E. B., Biswas, G., & Price, M. J. (2021). How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system? *Learning and Instruction, 72*, 101200. https://doi.org/10.1016/j.learninstruc.2019.04.001

Van der Stel, M., & Veenman, M. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learning and Individual Differences, 18*(1), 128–134. https://doi.org/10.1016/j.lindif.2007.08.003

Veenman, M. V. J., Van Hout-Wolters, Bernadette H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Veenman, M. V. J.,Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction, 14*(1), 89–109. https://doi.org/10.1016/j.learninstruc.2003.10.004

Veenman, M., & Van Cleef, D. (2018). Measuring metacognitive skills for mathematics: students' self-reports versus on-line assessment methods. *ZDM, 51*(4), 691-701. https://doi.org/10.1007/s11858-018-1006-5

Wetzel, E., Böhnke, J. R., & Brown, A. (2016). Response Biases. In F. T. L. Leong, D. Bartram, F. M. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment* (pp. 349-363). Oxford University Press.

Wolcott, M. D., & Lobczowski, N. G. (2021). Using cognitive interviews and think-aloud protocols to understand thought processes. *Currents in Pharmacy Teaching and Learning, 13*(2), 181-188. https://doi.org/10.1016/j.cptl.2020.09.005