RESEARCH ARTICLE   **OPEN ACCESS**

INTRODUCING RATIONALITY AND CONTENT VALIDITY OF SLAT-THINKING

*[1]Cristiano Mauro Assis Gomes, [1]Isabella Santos Linhares, [2]Enio Galinkin Jelihovschi
and [1]Marina Nogueira dos Santos Rodrigues

[1]Laboratório de Investigação da Arquitetura Cognitiva (LAICO), Universidade Federal de Minas Gerais, Brazil;
[2]Universidade Estadual de Santa Cruz (UESC), Ilhéus, Brazil

ARTICLE INFO

ABSTRACT

The students' learning approaches theory investigates a significant topic which is the interaction between subject and object of knowledge and its impact on learning. Nevertheless, the exclusive use of self-report instruments for its measures has become a fundamental limitation in that field. In this article, rationality and content validity of SLAT-Thinking (Students' Learning Approach Test) are introduced as the first test to measure learning approaches by means of performance. We also present its conceptual basis, building strategies and structure. The assessment of four construct experts, one expert in Portuguese and 10 people from the target-audience regarding the content validity, is shown. A new category was created to classify the items, the answer key of two items was changed and the statement of one item was reformulated. The experts certified the content validity of the test, and the target-audience stated the test was easy to understand and to perform.

INTRODUCTION

The theory of learning approaches has as its research focus the subject-object interaction of knowledge, in which it determines that this interaction is shaped by two basic approaches: deep approach and superficial approach (Gomes, 2013). The superficial approach indicates a poor interaction of the subject with the objects of knowledge. It is empirically identified in situations where the subject, when interacting with the object of knowledge, demonstrates poor understanding, poor identification of conceptual relationships, rote learning, memorization focused only on explicit information, and extrinsic motivation (Gomes, Golino, Pinheiro, Miranda & Soares, 2011). The deep approach, on the other hand, indicates an active and rich interaction of the subject with the objects of knowledge and can be verified empirically by interactions in which the subject demonstrates deep understanding, obtains explicit and implicit information, builds broad conceptual relationships, construes personal meaning, transfers learning to new contexts and is intrinsically motivated (Gomes, 2010). The theory postulates that the deep approach produces a better learning compared to the superficial approach (Gomes, 2011; Gomes & Golino, 2012). Strong evidence from two meta-analyzes corroborates this postulate, nevertheless it also indicates that the deep approach has a weak association with academic achievement. When selecting all existing studies at the time, where the instruments with alpha reliability below 0.50 had been removed, the Watkins (2001) meta-analysis investigated 60 correlations from university and school samples from 15 countries (N = 28.053). Their results showed that the deep approach has a correlation of .16 with academic performance, while the superficial approach has a correlation of -.11. He has also found that correlations obtained from school and university samples from Western and non-Western countries and from different instruments were similar. A second meta-analysis by Richardson, Abraham and Bond (2012) examined the English language studies from the PsycINFO and Web of Knowledge databases from 1997 to 2010, and selected 22 superficial approach studies (N = 5211) and 23 deep approach studies (N = 4838) which associated approaches with university global performance (GPA- *grade point average*). They found a correlation of .14 between the deep approach and GPA, with a 95% confidence interval ranging from .09 to .18, but a -.18

correlation between the superficial approach and GPA, with a 95% confidence interval ranging from -.25 to -.10. It is noteworthy that both meta-analyzes obtained similar results. In sum, the results ratify the theory of a positive correlation of the deep approach with academic performance, as well as a negative correlation between it and the superficial approach, suggesting that the deep approach is better than the superficial approach in the process of learning. However, the low correlation suggests that learning approaches, at least in the way they are evaluated, are not important for student's performance, a fact that is not emphasized by the theory. Tests based on performance, such as intelligence (Alves, Gomes, Martins, & Almeida, 2016, 2017, 2018; Golino & Gomes, 2019; Golino, Gomes & Andrade, 2014; Gomes, 2010a, 2010b, 2011a, 2012; Gomes & Borges, 2007, 2008, 2009a, 2009b; Gomes, de Araújo, Ferreira & Golino, 2014; Gomes & Golino, 2012a, 2012b, 2015; Gomes, Golino, Santos & Ferreira, 2014; Muniz, Gomes, & Pasian, 2016; Valentini et al., 2015) and metacognition (Cardoso, Seabra, Gomes, & Fonseca, 2019; Dias et al., 2015; Gomes & Golino, 2014; Gomes, Golino, & Menezes, 2014; Gomes, Golino, Santos, & Ferreira, 2014; Pereira, Golino, M. T. S., & Gomes, 2019; Reppold et al., 2015) seems to be much more important than students' approaches to learning to predict academic achievement. Furthermore, socioeconomic variables (Gomes, Amantes & Jelihovschi, 2020; Gomes & Jelihovschi, 2019; Gomes, Lemos, & Jelihovschi, 2020; Pazeto, Dias, Gomes & Seabra, 2019) are in a prominent place, as intelligence and metacognition, regarding the prediction of students' performance.

In its turn, motivational and self-reference variables, such as students' beliefs about the teaching-learning processes (Alves, Flores, Gomes & Golino, 2012; Gomes & Borges, 2008a), motivation for learning (Gomes & Gjikuria, 2018), academic self-reference (Costa, Gomes, & Fleith, 2017) and learning styles (Gomes, Marques, & Golino, 2014; Gomes & Marques, 2016) seems to be similar in importance in comparison to students' approaches to learning, concerning students' achievement (Gomes, 2010a, 2011a; Gomes, Golino, Pinheiro, Miranda, & Soares, 2011; Gomes & Golino, 2012b; Gomes, 2013). The students' learning approaches theory conceptually assumes that when there is an weak correlation between students' approaches to learning and students' achievement, it is because there is an improper educational assessment which reinforce inappropriate forms of teaching and learning. The field of learning approaches has two basic forms of deep and superficial approaches empirical identification. One is characterized by the phenomenographic qualitative method while the other is characterized by the measurement by means of self-report instruments. The qualitative method was elaborated and used mostly in the early days of the theory while the quantitative method, on the other hand, was incorporated since late 1970s onwards, becoming later on, the dominant procedure (Contreras, Salgado, Hernández-Pina, & Hernández, 2017). It should be noted that Marton and Säljö (1976) were the first to theoretically and empirically use the phenomenographic method to support learning approaches, categorizing students' performance in the reading context. In their categorization, they identified a level of superficial process of learning, where students had poor understanding of the text, a rote learning, as well as poor identification of conceptual relationships. The authors also identified a level of deep learning process where students have shown a deep understanding of the text, meaningful memorization, strategies

that maximized the identification of conceptual relationships, as well as identification of both explicit and implicit information in the text. At the same time, Biggs and collaborators in Australia, as well as Entwistle and collaborators in the United Kingdom, continued investigations into the approaches, but developed self-report questionnaires to quantitatively assess these constructs (Contreras et al., 2017; Watkins, 2001). The exclusive use of phenomenography and self-report instruments shows a number of problems. At its core, phenomenography depends on judges evaluating and categorizing students' performance in subject-to-object interaction tasks to empirically identify approaches. This process is complex because it demands the evaluation of judges, bringing with it the possibility of confirmatory bias, since the judge is himself the creator of the categories, and might analyze the performance of students with a conscious or unconscious bias of his or her own expectations or previous assumptions. Even with all methodological care and high correlation among judges' judgments, the categories created are not free of relevant confirmatory bias. As stated by Das-Smaal (1990), "... real-world features, objects, and events may be categorized into an infinite number of forms. In addition, our perception is highly selective and therefore biased" (p. 349).

The reader interested in a detailed description of categorization biases and how they articulate with the confirmatory bias may refer to Das-Smaal's (1990) work. Phenomenography, in its need for judge evaluation, makes the process complex and difficult to apply in large samples while the quantitative self-report method, on the other hand, allows the study of a large number of people. However, self-report instruments measure their target constructs through people's perception or judgment about these constructs, which tends to generate considerable response bias, as shown by a large body of research. The interested reader can refer to the work of Wetzel, Böhnke and Brown (2016). In addition to the response bias, accuracy is also a problem, because self-report instruments may generate very inaccurate and noisy scores if respondents do not have a good understanding of their own internal process as well as a good comprehension of the target construct. Strong evidences have shown that learning approaches and school performance have a low correlation so that considering the fact that the approaches have been measured exclusively by self-report instruments, it is plausible to suppose that these evidences, despite being robust, are mistaken, because they may be based on information with strong noise. This assumption is different from that commonly held by field theorists that the low correlation found is caused primarily by the educational system itself since it does not produce or evaluate high quality learning (Watkins, 2001). Of course, we do not rule out this explanation, but we also consider that the evidences may be wrong.

Taking all that into account, this article presents to the scientific community the first test that aims to evaluate students' learning approaches through student performance: the Students' Learning Approach Test - Identification of Thinking Contained in Texts (SLAT-Thinking). SLAT-Thinking is able to integrate the best of the methodologies applied in the field of learning approaches. It empirically identifies approaches through students' performance in subject-to-object interaction tasks, as occurs in phenomenography, as well as data collection in large samples, such as self-report instruments. Moreover, its measures do not

generate strong response and confirmatory bias, since it does not depend on judges and either the perception or judgment of the respondents. In this article, we present in detail the rationality of SLAT-Thinking together with the first evidence of validity of this test, through content validity analysis.

Introducing the rationality of SLAT-Thinking: conceptual basis, set up and structure strategies

The SLAT-Thinking is based on the basic premise that the learning approach can be measured by the performance of people in tasks that require the interaction subject-object of knowledge. This premise is in line with the theory. The deep approach is defined by the theory of learning approaches as the active posture of the subject in relation to the objects of knowledge. It is operationally delimited through a series of subject-specific behaviors when interacting with knowledge objects, such as knowing how to obtain both literal and implicit information when interacting with the knowledge object, how to transfer learned knowledge to other pertinent areas, form meaningful relationships or build understanding and personal meaning about what is being learned. In turn, the superficial approach is operationally delimited through behaviors that represent a poor interaction between the subject and the objects of knowledge, that is, not knowing how to obtain literal or implicit information, not knowing how to transfer the knowledge learned or forming very low meaningful and poor relationships. SLAT-Thinking is based on a complementary premise, besides the basic one, that each behavior that carries out learning approaches is better put in motion depending on the type of task that involves the subject-object interaction. For example, some subject-object interaction tasks may be more appropriate to mobilize the behavior of transferring learned knowledge to other areas, while other tasks may be more appropriate to focus on the behavior of knowing how to obtain explicit and implicit information present in the object of knowledge.

Given the large number of behaviors that work out the learning approach construct, and considering the complementary premise that each behavior is best mobilized by specific tasks, there is no appeal that a single test should seek to mobilize too many behaviors. This may be unfeasible or make the test too long. SLAT-Thinking considers this issue and seeks to measure people's approach in identifying the author's thinking in a given text. This choice is not random. We are constantly stating our way of thinking, either through speech or written texts. In our words we show our thoughts, beliefs, and values on a particular theme, as well as opening ourselves to dialogue with others, or closing ourselves in our own opinions. When uttering our thinking, we also hear or read about the thoughts of others so that we can influence or be influenced. Besides, we may also manipulate or be manipulated by intentionally designed thought statements to deceive, mislead and manipulate. Finally, the process of utterance of thought and dialogue is complex and has many functions and possibilities. In a globalized society where communication technologies play an increasingly prominent role in people's lives and in the formation and proliferation of ideas and actions, correctly understanding the kind of thought contained in a speech or text is a fundamental behavior. It is crucial for the shaping of a critical citizen. Through it, the citizen can correctly identify the argumentative structure of his own thinking and that of other people, so that he or she becomes able to understand in depth the thought contained in

the different speeches and texts, making him or her more apt to agree or disagree, to criticize, to make choices, and propose new thoughts. Moreover, throughout this behavior, the citizen becomes less inclined to mix up his own thinking and prior knowledge with the thinking presented by other people, just as he or she becomes less inclined to analyze the thought contained in a speech or text through a superficial analysis based on isolated fragments or terms. By taking into account the theory of learning approaches, it is assumed that the act of correctly identify an author's thinking contained in a given text is a very active interaction of the subject and is therefore, a good operational indicator of the deep approach in the 21st century. In addition to the basic and complementary assumptions, SLAT-Thinking has an operational premise that defines prior knowledge and logical analysis of arguments as the central elements to identify the author's way of thinking contained in a text. This premise considers that without prior knowledge the occurrence of reading is not possible. For example, when reading a text about "a walk in the park", the reader needs to have a minimum prior knowledge of what is walk or park. The premise also considers that, despite of being a requirement, prior knowledge needs to be well used so that the author's way of thinking contained in a text is correctly identified. Misuse of prior knowledge can lead the reader to mistakenly assume that a text has certain thoughts, values or beliefs, while in reality they are merely a misguided projection of his or her own prior knowledge to the text.

The logical analysis of the arguments provides for a correct identification of the argumentative structure presented by the author. This analysis also provides a precise scrutiny of both textual elements and their connections, such as the reviewing of gaps, imperfections, misrepresented arguments, among other aspects that might stand for the way the author organizes and exposes his thinking. The operational premise also holds that, by definition, the relationship between prior knowledge and logical analysis is indivisible. Improper use of prior knowledge is only possible if the reader also makes an incorrect logical analysis of the arguments, as a correct logical analysis does not allow the reader to confuse his prior knowledge with the author's thinking. All logical analysis in argumentative texts demands prior knowledge about each term presented. Through the operational premise, it is assumed that a high performance in SLAT-Thinking indicates a respondent who uses his previous knowledge to understand the terms in the text, but differentiates the author's thinking from his own way of thinking, by logically analyzing the arguments contained in the text. Poor performance in SLAT-Thinking, on the other hand, indicates a respondent who possibly does not adequately differentiate between his own way of thinking and the author's way and does not logically analyze the author's arguments contained in the text. This respondent does not correctly understand how textual terms are connected by the author himself and, at the same time, mistakenly infers that the author makes a series of connections that are really only part of his own prior knowledge. At its worst, it is as if this respondent reads with his eyes closed, as he basically "reads" his own prior knowledge of the subject rather than reading the author's arguments. As SLAT-Thinking measures people's approach in identifying the author's thinking in a given text, it is also assumed that other tests may and should be designed to have their focuses on other behaviors with the function of putting the learning approaches into operation. This will allow, in the future, a wide range of tests in order to address different behaviors that represent the diversity of aspects that

circumscribe learning approaches, bringing considerable benefit to the field of measurement. SLAT-Thinking is structurally divided in two texts and 12 items related to each text. Each item is composed by a statement that may or may not represent the author's thinking in a given text. The respondent has to read the statements contained in each item and identify whether or not this statement represents the author's thinking. To answer each item, the respondent must consider three options and mark one of them. If the respondent believes that the item's statement represents the author's thinking, then he or she should mark the R (represented) column of the answer options. If the respondent understands that the item does not represent the author's thinking, then he or she should mark the N (not represented) column of the answer options. If the respondent believes that it is not possible to answer whether or not the item represents the author's thinking contained in that text because it did not provide enough information, then he or she should mark an X in the Z (no answer) column of the answer options.

Both SLAT-Thinking texts have a very similar structure, with a total of eight sentences, one assumption, two arguments directly related to the assumption, and one sentence containing the important terms of items statements. Both texts have roughly the same number of words. The first text has 252 words, 106 referring to the first argument, 61 to the second argument, and 85 words to the sentence containing the important terms of the items statements. The second text has 225 words, 98 referring to the first argument, 47 to the second argument and 80 words to the sentence containing the important terms of the items statements. The texts always start with the assumption in the first sentence, presenting the main idea of the author. The second sentence shows an argument, that is, a statement that seeks to support the assumption. The third sentence presents a scientific information regarding the argument of the second sentence and it also presents a cause for the argument described in the second sentence. The fourth and fifth sentences add new elements (associative or causal) about the causal relationship presented in the third sentence. The sixth sentence shows the second argument that directly supports the assumption and the seventh sentence presents a scientific information related to the second argument. Finally, the eighth sentence only states the importance of the terms contained in the statement of the items.

The content of the texts was strategically designed so that the previous knowledge of the respondent be intensively activated. The test's authors designed both texts with a view to highlight controversial topics of strong social relevance that are currently being discussed in modern society. The first theme involves violence against women and the second concerns prejudice against homosexual couples. In theory, controversial topics of strong social relevance provide a rich context to verify whether the respondent is able to correctly identify the structure of author's thinking in a given text. The assumption of each of the two texts has a strong common sense character, when referring to well-established previous knowledge, that is, opinions already well established by most readers. In other words, the assumptions of the two texts follow a current, politically correct, and pro-minority thinking. The assumption of the first text states that "Women have constantly been victims of violence," whereas the assumption SLAT of the second text states that "Even nowadays there is a strong prejudice against homosexual couples." The arguments of each text follow the assumption in order to support it. They do not necessarily oppose common sense values and thoughts about the theme involved in each text, but develop specific ideas of their own to support the assumption. In this context, the reader who does not correctly analyze the logic of the text's arguments, and also does not have a firm attitude and engagement to accurately identify the author's thinking and differentiate it from his own or other people's thinking, have a tendency to interpret the text as a series of statements or a series of thoughts when this is actually not true. The last sentence of the texts was also conceived for strategic reasons. It would be possible for a respondent with a strong superficial approach, to answer certain items as not representative of the author's thinking, just because certain terms of the statements of these items were not present in the text. In this case, the respondent, through the superficial approach, could get right answers in certain items, what would generate a serious problem, since the test was designed to assign positive scores to the deep approach. The last sentence of the texts seeks to prevent this from happening by containing the essential terms of the statements of the items.

The items' arguments were constructed by taking five categories as reference:

1. **Correct items:** These items are designated as correct, whenever their statements express the author's thinking in the text. The arguments of these items are not literal sentences taken from the text itself. This prevents the reader from correctly identifying that the item's statement expresses the author's thinking only by the similarity of the sentences.

2. **Incorrect and minority sensitive items:** These items are called incorrect because their statements do not express the author's thinking in the text. They also represent favorable, minority-sensitive thoughts. To the extent that the assumption (core idea) of the two texts of the test is strongly in line with a politically correct and minority-friendly stance, it is expected that the respondent who does not logically analyze the arguments of the text itself and also does not have firm commitment in differentiating the author's thinking from other thoughts on the same subject, including his own, may mistakenly consider that items in this category also represent the author's thinking.

3. **Incorrect and no sensitive to minority items:** As with incorrect and minority-sensitive items, items in the third category also do not express the author's thinking in the text and are therefore called incorrect. Besides of being incorrect, they represent thoughts that are not sensitive to minorities, that is, they express thoughts that are not favorable to minority struggles or ideologies. It is possible that the respondent mistakenly concludes that items in this category represent the author's thinking if the statements contained in these items represent or activate some thought or belief coming from the respondent's own prior knowledge.

4. **Incongruent items:** These items are statements that have nothing to do with the subject or arguments of the text. It is assumed that if the respondent erroneously concludes that this type of item represents the author's thinking, it is likely that he is being conducted by a strong superficial approach to the extent that either the actual non-reading of the texts themselves or the random response in relation to them is an intense form of superficial approach. The respondent may also make the mistake of concluding that it is not possible to answer (answer option Z) whether or

not these items represent the author's thinking. In this case, it is possible that the respondent mistakenly deems the analysis and identification of the author's thinking as entailing the understanding of the author's thinking beyond the analyzed text. This way of thinking is misleading because it completely evades the test instructions. Moreover, in this line of reasoning, it would not be possible to answer any of the test items, as the reader would not have access to the author's thinking that goes beyond that text. In this way, the arguments presented in the text would not be sufficient to provide any conclusion about whether or not the statements of the items actually represent the author's thinking.

5. **Incorrect and neutral items:** These items also do not express the author's thinking and are therefore called incorrect. Since they bring only quantitative or qualitative information that does not represent clearly favorable or unfavorable positions to minorities, their struggles and ideologies, these items are called neutral. It should be highlighted that this category was created a posteriori, based on the suggestions of the reviewers of the content validity analysis of SLAT-Thinking. Considering the characteristics of this group of items, it is possible, among other possibilities, that the respondent mistakenly concludes that they represent the author's thinking, since the information provided in the statements of these items either intermingle with the respondent's own beliefs and thoughts or appear to represent logically the arguments of the text.

Both texts have a similar amount of items for each category. For example, if text 1 has five items in category 2, text 2 will tend to have an equal or close number of items in this category.

MATERIALS AND METHODS

Participants: The content validity analysis of SLAT-Thinking involved three groups of participants. The first consisted of four experts in the construct of students' learning approaches, the second consisted of a Portuguese language expert, and the third consisted of 10 college students or graduates representing the test target audience. The group of specialists in the construct consisted of three psychologists from different fields (clinical, assessment and educational psychology) and a physical therapist, aged between 30 and 45 years, mostly female attending a postgraduate program (masters or doctorate). The Portuguese language specialist was 50 years old female, graduated in Letters. All participants in the target population were adults (half female), aged 19 to 50 years old, trained in various fields of knowledge (Law, Control and Automation Engineering, Pedagogy, Law, Civil Engineering, Dentistry and Psychology).

Instrument

SLAT-Thinking: SLAT-Thinking was developed by Cristiano Mauro Assis Gomes and Isabella Santos Linhares, in Portuguese, in 2018, at the Cognitive Architecture Mapping Laboratory (*Laboratório de Investigação da Arquitetura Cognitiva* – LAICO), with the aim, as far as we know, to be the first test to gauge learning approaches through respondents' performance. The test assesses people's approach in identifying an author's thinking in a given text and is aimed at people who at least have not completed senior high school.

The test has two texts and 12 items related to each text. Each item is formed by a statement that may or may not represent the author's thinking. The respondent's task is to read each text, read the statement in each item, and identify whether or not that statement represents the author's thinking contained in the text. In order to answer each item, the respondent must consider three response options: R (represents the author's thought contained in the text), N (does not represent the author's thought contained in the text) and Z (there is no way to answer). As for the SLAT-Thinking measure, a higher score implies a greater deep approach of the respondent to identify the author's thinking, while a lower score implies a greater superficial approach. Since the test has 24 items, SLAT-Thinking has a minimum raw score of 0 points and a maximum raw score of 24 points.

Data collection and analysis procedures: Table 1 represents the content validity steps and schematically shows the tasks requested for each sample (study participants), as well as the data collection and analysis strategies. All the procedures followed the ethical guidelines for research. This study is part of a research approved by a Research Ethics Committee of Brazil. The construct specialists got the SLAT-Thinking and the assessment protocol, and were asked to answer the test items and then the protocol. This required the expert to classify each of the items into one of four categories (1. correct items, 2. incorrect and minority-sensitive items, 3. incorrect and non-minority-sensitive items, 4. incongruent items) and give the underlying reasons of her or his rating. The expert should also judge whether each item was appropriate to represent the category she or he classified as related to the item. If an item was assessed as inappropriate, the expert should write a justification of the reasons behind his or her judgment. In this evaluation process, no information was given to the experts on how the authors of the SLAT-Thinking classified each item into the given categories. The expert assessments were then analyzed by the researchers and grouped into categories.

Two distinct categorizations were rendered. The first of them aimed to create categories that could represent different types of agreement, relating the ratings of the experts and the authors of the test. The second categorization involved forming categories that could represent different types of reasons related to the judgments of the experts regarding the adequacy of the items. The experts were subsequently invited to attend weekly meetings with the researchers to collectively discuss the evaluation of each item, clarify controversies and make suggestions. At these meetings, the researchers presented the experts a logical analysis of the statement of each item and the statements of the texts, in order to logically support the classification formulated by the authors of the test. At the same time, the categorizations created by the researchers, regarding the expert evaluations, were also presented and discussed. The experts should assess whether or not they agreed with the logical analysis presented, as well as whether the categorizations of the researchers correctly represented their classification process and judgment of the appropriateness of the items. If any expert disagreed with any element presented, then she or he should present the arguments of his disagreement. In this case, there was a collective discussion and, if necessary, changes to the test were made. The Portuguese specialist was asked to take the test, in the presence of one of the researchers, and to review aloud the wording of the instructions, texts and statements of

Table 1. SLAT-Thinking Content Validity Steps

Sample	Task	Collection	Analysis	Step
Construct experts	- Answer the SLAT-Thinking. - Respond to the Evaluation Protocol for the purpose of: 1. Classify each item into one of four categories. 2. Assess if the item is appropriate to represent the chosen category.	- Referral to experts by email of: 1. SLAT-Thinking 2. Evaluation protocol to classify items and assess their suitability.	- Building categories, by the researchers, to evaluate: 1. The reasons and the degree of agreement among the experts regarding the classification of the items. 2. The reasons related to the judgment of the adequacy of the items.	Step 1
	- Attend meetings with researchers to collectively discuss the evaluation of each item.	- Weekly in-person meetings. - Keeping the record of expert comments at the meetings.	- Examination of the logical relationships contained in the texts and statements of each item, by comparing the classification of the items made by the experts and the authors of the test.	
Portuguese Specialist	- Answer the SLAT-Thinking and evaluate aloud the wording of the instructions, texts and items of the instrument. - Logically analyze the arguments of the texts and items and evaluate the answer key of the test.	- In-person interview with the specialist. - Keeping the record of expert comments during the interview.	- Checking errors pointed out by the expert.	Step 2
Target Audience	- Answer the SLAT-Thinking.	- Personal and individual application of SLAT-Thinking, followed by an interview. - Recording the comments of each person during the interviews.	- Test understanding check. - Verifying the reasons underlying the response of each item.	Step 3

the items. Then, the answer key of the test was presented to the expert, who should evaluate its suitability. During the expert assessment process, all of their suggestions regarding the content of the test were recorded. Each participant of the target audience answered the test individually and then participated in an interview with one of the researchers. During the interviews, a semantic analysis of the instrument was made, assessing the clarity of the test instructions and their feasibility.

RESULTS AND DISCUSSION

Since the content validity analysis involved three distinct steps (Table 1), the results will follow this same sequence.

Construct specialists: categorization regarding the item classification task: The researchers elaborated six categories in order to understand the classification made by the experts and their degree of similarity to the classification conceived by the authors of the test. Except for the first and last of the categories, it will not be possible to show the number of items related to them, or otherwise the answer key of the test may be shown, even if implicitly. The Consensus category represents a pattern in which all experts and test authors identically classified the items. This pattern occurred in nine of the 24 test items. The Strong Representation Bias category represents the context in which all experts classified certain items as representatives of the author's thinking, while the test authors classified these items as non-representatives. Throughout collective meetings, it has been concluded that an item in this category was misclassified by the test authors and that the answer key should be changed. Regarding the other items in this category, it was found that the experts' classification was incorrect, due to logical errors made by them in interpreting the statements of the items or texts. The Weak Representation Bias category indicates a pattern of strong agreement in which three of the four experts, as well as the authors of the test, classified certain items as not representative of the author's thinking. In the collective meetings, it has been found that the classifications that represented the exceptions exhibited logical errors in the analysis and interpretation of the statements of the

meetings, it has been found that the classifications different from those made by the test authors involved an inadequate interpretation of the items and texts. In a final review, following meetings with the experts, it has been found that an item in this category should be changed from representative to non-representative of the author's thinking.

Construct experts: categorization regarding the judgment of item appropriateness: The researchers elaborated five categories in order to understand the reasons behind the experts' judgment of the adequacy of the items. Except for the first and last of the categories, it will not be possible to show the number of items related to them, or otherwise the answer key of the test may be shown, even if implicitly. The first category, Total Suitability, indicates a pattern of total expert agreement on the suitability of certain items. This category covered most of items (15 out of 24 items). The second category, Sensitivity, represents the context in which some experts considered certain items inappropriate as they had doubts whether these items could be classified as sensitive or not sensitive to minorities. The perceived inadequacy was solved by creating the category incorrect and neutral items. The Non-Representation Error category is similar to the second category, in the sense that some experts considered certain items inappropriate, by understanding that these items do not represent the authors' thinking and could not be classified as sensitive or not sensitive to minorities. However, throughout collective meetings, it has been found that the judgment of inadequacy of those items was specifically related to misclassification of the experts themselves. The fourth category, Representation Error, represents the context in which some experts understood that certain items represent the authors' thinking but were poorly written. However, throughout the collective meetings, it has been found that the judgment of inadequacy of the items, was specifically related to the experts' errors of understanding about certain excerpts contained in the statements of these items. Finally, the last category, Representation Doubt, is quite similar to the previous one, as some experts also thought that certain items were poorly written. Unlike the previous category, experts were not sure of their classification, attributing this insecurity to the inadequate writing of these items. Also as in the

previous category, this judgment was caused by misunderstandings of experts on certain portions of the wording of these items.

Portuguese expert: The opinion of the Portuguese language expert was in complete agreement with the test's answer key. She also pointed out a few grammatical errors in the instrument that were analyzed and corrected.

Target audience: The target audience favorably evaluated the instrument. The test instructions were indicated, by semantic analysis, as understandable and feasible, and participants errors were specifically caused by inherent difficulties to the challenge of the test itself.

Conclusion

This article exhibits original and relevant contributions to the field of learning approach studies. The first contribution is a new presentation of the problem of learning approaches measurement and its implications. We show that the field concentrates all its measures on self-report instruments, highlighting the fact that the absence of tests that measure the approaches through respondents' performance is an important limitation in the area. It is possible that solid evidence about the weak correlation between approaches and academic achievement is merely a symptom of the exclusive use of self-report tools to gauge the approaches. In turn, the second contribution of this paper involves the proposition and detailed presentation of a test of learning approaches through people's performance: the SLAT-Thinking. The third contribution is the presentation of initial evidence of SLAT-Thinking validity. A fourth contribution involves the opening of a new research agenda, not possible before due to the lack of performance tests in the area. Although promising, the contributions of this article have limitations since only initial evidence of SLAT-Thinking validity are presented, bringing the need for further research in the field which should be directed to the search for evidence regarding the structural validity and external validity of SLAT-Thinking. The initial evidence in this article only points to the beginning, more robust evidences should be brought forward as research in the field continues.

REFERENCES

Alves, A. F., Gomes, C. M. A., Martins, A., & Almeida, L. S. (2016). Social and cultural contexts change but intelligence persists as incisive to explain children's academic achievement. *PONTE: International Scientific Researches Journal, 72*(9), 70-89. doi: 10.21506/j.ponte.2016.9.6

Alves, A. F., Gomes, C. M. A., Martins, A., & Almeida, L. S. (2017). Cognitive performance and academic achievement: How do family and school converge? *European Journal of Education and Psychology, 10*(2), 49-56. doi: 10.1016/j.ejeps.2017.07.001

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

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. *Revista E-PSI, 1*, 97-117. Retrieved from https://revistaepsi.com/artigo/2012-ano2-volume1-artigo5/

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. doi: 10.3389/fpsyg.2019.01784

Contreras, M. S., Salgado, F. C., Hernández-Pina, F., & Hernández, F. M. (2017). Enfoques de aprendizaje y enfoques de enseñanza: Origen y evolución. *Educación y Educadores*, 20(1), 65-88. DOI: 10.5294/edu.2017.20.1.4

Das-Smaal, E. A. (1990). Biases in categorization. In J-P. Cavearni, J-M. Fabre, & M. Gonzalez (Eds.), *Advances in Psychology: Cognitive Biases*, volume 68 (pp. 349-387). North-Holland: Elsevier Science.

Dias, N. M., Gomes, C. M. A., Reppold, C. T., Fioravanti-Bastos, A., C., M., Pires, E. U., Carreiro, L. R. R., & Seabra, A. G. (2015). Investigação da estrutura e composição das funções executivas: análise de modelos teóricos. *Psicologia: teoria e prática*, 17(2), 140-152. doi: 10.15348/1980-6906/psicologia.v17n2p140-152

Golino, H. F., & Gomes, C. M. A. (2014a). Four Machine Learning methods to predict academic achievement of college students: a comparison study. *Revista E-Psi, 1*, 68-101. Retrieved from https://revistaepsi.com/artigo/2014-ano4-volume1-artigo4/

Golino, H. F., & Gomes, C. M. A. (2019) *TDRI: Teste de Desenvolvimento do Raciocínio Indutivo.* São Paulo: Hogrefe.

Golino, H. F., Gomes, C. M. A., & Andrade, D. (2014). Predicting academic achievement of high-school students using machine learning. *Psychology, 5,* 2046-2057. doi:10.4236/psych.2014.518207

Golino, H. F., Gomes, C. M. A., Amantes, A., & Coelho, G. (2015). *Psicometria Contemporânea: compreendendo os Modelos Rasch.* São Paulo: Casa do Psicólogo.

Golino, H. F., Gomes. C. M. A., Commons, M. L., & Miller, P. M. (2014). The construction and validation of a developmental test for stage identification: Two exploratory studies. *Behavioral Development Bulletin, 19*(3), 37-54. doi: 10.1037/h0100589

Golino, H.F., & Gomes, C. M. A. (2014b). 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), p.e6. doi:10.5334/jopd.af

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

Gomes, C. M. A. (2007). Softwares educacionais podem ser instrumentos psicológicos. *Psicologia Escolar e Educacional, 11*(2), 391-401. doi: 10.1590/S1413-85572007000200016

Gomes, C. M. A. (2010a). Avaliando a avaliação escolar: notas escolares e inteligência fluida. *Psicologia em Estudo, 15*(4), 841-849. Retrieved from http://www.redalyc.org/articulo.oa?id=287123084020

Gomes, C. M. A. (2010b). Estrutura fatorial da Bateria de Fatores Cognitivos de Alta-Ordem (BaFaCalo). *Avaliação Psicológica, 9*(3), 449-459. Retrieved from http://pepsic.

bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000300011&lng=pt.

Gomes, C. M. A. (2010c). Perfis de Estudantes e a relação entre abordagens de aprendizagem e rendimento Escolar. *Psico (PUCRS. Online)*, *41*(4), 503-509. Retrieved from http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/6336

Gomes, C. M. A. (2011a). Abordagem profunda e abordagem superficial à aprendizagem: diferentes perspectivas do rendimento escolar. *Psicologia: Reflexão e Crítica*, *24*(3), 438-447. doi: 10.1590/S0102-79722011000300004

Gomes, C. M. A. (2011b). Validade do conjunto de testes da habilidade de memória de curto-prazo (CTMC). *Estudos de Psicologia (Natal)*, *16*(3), 235-242. doi: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). *Gerais: Revista Interinstitucional de Psicologia*, *5*(2), 294-316. Retrieved from http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1983-82202012000200009&lng=pt&tlng=pt.

Gomes, C. M. A. (2013). A construção de uma medida em abordagens de aprendizagem. *Psico*, *44*(2), 193-203. Available from http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/viewFile/11371/9638

Gomes, C. M. A., & Borges, O. (2008b). Qualidades psicométricas de um conjunto de 45 testes cognitivos. *Fractal: Revista de Psicologia*, *20*(1), 195-207. doi:10.1590/S1984-02922008000100019

Gomes, C. M. A., & Borges, O. (2009c). Qualidades psicométricas do conjunto de testes de inteligência fluida. *Avaliação Psicológica*, *8*(1), 17-32. Retrieved from http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712009000100003&lng=pt&tlng=pt.

Gomes, C. M. A., & Borges, O. N. (2007). Validação do modelo de inteligência de Carroll em uma amostra brasileira. *Avaliação Psicológica*, *6*(2), 167-179. Retrieved from http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712007000200007&lng=en&tlng=pt.

Gomes, C. M. A., & Borges, O. N. (2008a). Avaliação da validade e fidedignidade do instrumento crenças de estudantes sobre ensino-aprendizagem (CrEA). *Ciências & Cognição (UFRJ)*, *13*(3), 37-50. Retrieved from http://www.cienciasecognicao.org/revista/index.php/cec/article/view/60

Gomes, C. M. A., & Borges, O. N. (2009a). O ENEM é uma avaliação educacional construtivista? Um estudo de validade de construto. *Estudos em Avaliação Educacional*, *20*(42), 73-88. doi: 10.18222/eae204220092060

Gomes, C. M. A., & Gjikuria, E. (2018). Structural Validity of the School Aspirations Questionnaire (SAQ). *Psicologia: Teoria e Pesquisa*, *34*, e3438. doi:10.1590/0102.3772e3438

Gomes, C. M. A., & Golino, H. F. (2012a). O que a inteligência prediz: diferenças individuais ou diferenças no desenvolvimento acadêmico? *Psicologia: teoria e prática*, *14*(1), 126-139. Retrieved from http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872012000100010&lng=pt&tlng=pt.

Gomes, C. M. A., & Golino, H. F. (2012b). Validade incremental da Escala de Abordagens de Aprendizagem (EABAP). *Psicologia: Reflexão e Crítica*, *25*(4), 400-410. doi:10.1590/S0102-79722012000400001

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

Gomes, C. M. A., & Golino, H. F. (2015). Factor retention in the intra-individual approach: Proposition of a triangulation strategy. *Avaliação Psicológica*, *14*(2), 273-279. doi: 10.15689/ap.2015.1402.12

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*. doi: 10.1080/1743727X.2019.1654992

Gomes, C. M. A., & Marques, E. L. L. (2016). Evidências de validade dos estilos de pensamento executivo, legislativo e judiciário. *Avaliação Psicológica*, *15*(3), 327-336. doi: 10.15689/ap.2016.1503.05

Gomes, C. M. A., de Araújo, J., Ferreira, M. G., & Golino, H. F. (2014). The validity of the Cattel-Horn-Carroll model on the intraindividual approach. *Behavioral Development Bulletin*, *19*(4), 22-30. doi: 10.1037/h0101078

Gomes, C. M. A., Golino, H. F., & Menezes, I. G. (2014). Predicting School Achievement Rather than Intelligence: Does Metacognition Matter? *Psychology, 5*, 1095-1110. doi:10.4236/psych.2014.59122

Gomes, C. M. A., Golino, H. F., Pinheiro, C. A. R., Miranda, G. R., & Soares, J. (2011). Validação da Escala de Abordagens de Aprendizagem (EABAP) em uma amostra brasileira. *Psicologia: Reflexão e Crítica, 24*(1), 19-27. doi: 10.1590/S0102-79722011000100004

Gomes, C. M. A., Golino, H. F., Santos, M. T., & Ferreira, M. G., (2014). Formal-Logic Development Program: Effects on Fluid Intelligence and on Inductive Reasoning Stages. *British Journal of Education, Society & Behavioural Science, 4*(9), 1234-1248. Retrieved from http://www.sciencedomain.org/review-history.php?iid=488&id=21&aid=4724

Gomes, C. M. A., Marques, E. L. L., & Golino, H. F. (2014). Validade Incremental dos Estilos Legislativo, Executivo e Judiciário em Relação ao Rendimento Escolar. *Revista E-Psi, 2*, 31-46. Retrieved from https://revistaepsi.com/artigo/2013-2014-ano3-volume2-artigo3/

Gomes, C. M. A.s, & Borges, O. N. (2009b). Propriedades psicométricas do conjunto de testes da habilidade visuo espacial. *PsicoUSF*, *14*(1), 19-34. Retrieved from http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-82712009000100004&lng=pt&tlng=pt.

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

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I – Outcome and Process. *British Journal of Educational Psychology*, *46*, 4-11. DOI: 10.1111/j.2044-8279.1976.tb02980.x

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

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.

Estudos de Psicologia, 24(3), 225-236. doi: 10.22491/1678-4669.20190024

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. *European Journal of Education Studies, 6*(7), 35-52. doi: 10.5281/zenodo.3477577

Pires, A. A. M., & Gomes, C. M. A. (2017). Three mistaken procedures in the elaboration of school exams: explicitness and discussion. *PONTE International Scientific Researches Journal, 73*(3), 1-14. doi: 10.21506/j.ponte.2017.3.1

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. doi:10.5281/zenodo.2313538.

Reppold, C. T., Gomes, C. M. A., Seabra, A. G., Muniz, M., Valentini, F., & Laros, J.A. (2015). Contribuições da psicometria para os estudos em neuropsicologia cognitiva. *Psicologia: teoria e prática, 17*(2), 94-106. doi: 10.15348/1980-6906/psicologia.v17n2p94-106

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and metaanalysis. *Psychol. Bull, 138* (2), 353–387. doi: 10.1037/a0026838.

Valentini, F., Gomes, C. M. A., Muniz, M., Mecca, T. P., Laros, J. A., & Andrade, J. M. (2015). Confiabilidade dos índices fatoriais da Wais-III adaptada para a população brasileira. *Psicologia: teoria e prática, 17*(2), 123-139. doi: 10.15348/1980-6906/psicologia.v17n2p123-139

Watkins, D. (2001). Correlates of Approaches to Learning: A Cross-Cultural Meta-Analysis. In R. J. Sternberg & L. F. Zhang (Eds.), *Perspectives on thinking, learning and cognitive styles* (pp. 132–157). Mahwah, NJ: Lawrence Erlbaum Associates.

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). New York, NY: Oxford University Press.

****