Predictive and incremental validity of Students' Learning Approach Test (SLAT-Thinking)

Cristiano Mauro Assis Gomes [1][a], **Jhonys de Araujo** [a] , **& Enio Galinkin Jelihovschi** [b] [2]

Universidade Federal de Minas Gerais, Belo Horizonte, Brasil [a]*; Universidade Estadual de Santa Cruz, Ilhéus, Brasil* [b].

ABSTRACT

SLAT-Thinking is the only test that evaluates and distinguishes stages of approaches through performance. Although SLAT-Thinking shows evidence of internal validity, its external validity has not yet been examined. In this paper we study the predictive and incremental validities of SLAT-Thinking. Two models were tested. The predictors were inductive reasoning, SLAT-Thinking approaches and Learning Approaches Scale (EABAP) approaches. The outcome was the Brazilian large-scale exam that evaluates the students that finish secondary education. In both models, the superficial approach of SLAT-Thinking was the main predictor, followed by inductive reasoning. The deep and superficial approaches of SLAT-Thinking were positively associated with academic achievement. While the deep intermediate approach was negatively associated to outcome. Non-linear relationships (positive and negative associations) were found in the two EABAP approaches and inductive reasoning with the outcome. This study shows evidence of predictive and incremental validity of SLAT-Thinking.

Keywords
students' approaches to learning, external validity, predictive validity, incremental validity

RESUMO

O TAP-Pensamento é o único teste que avalia estágios distintos de abordagens de aprendizagem por meio da performance do respondente. Embora este teste apresente evidências de validade interna, sua validade externa ainda não foi examinada. Este artigo investiga a validade preditiva e incremental do TAP-Pensamento. Dois modelos foram testados. Os preditores são o raciocínio indutivo e as abordagens do TAP-Pensamento e da Escala de Abordagens de Aprendizagem (EABAP). O desfecho utilizado foi o exame de larga-escala brasileiro que avalia os estudantes que finalizam a educação secundária. Em ambos os modelos, a abordagem superficial do TAP-Pensamento foi o principal preditor, seguida do raciocínio indutivo. As abordagens profunda e superficial do TAP-Pensamento associaram-se positivamente ao desempenho acadêmico, enquanto a abordagem intermediária-profunda associou-se negativamente ao desfecho. Relações não-lineares foram encontradas nas duas abordagens da EABAP e no raciocínio indutivo em relação ao desfecho. Este estudo traz evidências de validade preditiva e incremental do TAP-Pensamento.

Palavras Chave
abordagens de aprendizagem, validade externa, validade preditiva, validade incremental

---

[1] Correspondence about this article should be addressed **Cristiano Mauro Assis Gomes:** cristianomaurogomes@gmail.com

Validade Preditiva e Incremental do TAP-Pensamento

Introduction

The Learning Approach Test: Identification of Thought Contained in Texts (SLAT-Thinking) is an instrument developed to measure learning approaches. There are two reasons by which this instrument has shown itself to be very promising. While the other instruments of approaches are self-report tests, SLAT-Thinking is a performance test. Since self-report measures are permeated by biases, such as social desirability and acquiescence, SLAT-Thinking is advantageous to measure approaches, as it overcomes these biases (Gomes et al., 2020). In addition, traditional instruments measure approaches in terms of "all or nothing", that is, without identifying intermediate stages. In contrast, SLAT-Thinking allows the identification of intermediate stages, enabling the measurement of levels of development of the approaches (Gomes et al., 2020).

SLAT-Thinking shows evidence of internal validity. The content validity was examined and the judges attested the content validity of SLAT-Thinking, as well as, an expert in Portuguese approved the writing of the test and 10 people from the target audience certified that both the instructions and the task are easy to understand and execute (Gomes & Linhares, 2018). The structural validity of SLAT-Thinking was investigated through a sample of 622 higher education students from public and private institutions in the areas of biological, exact and human sciences (Gomes et al., 2020). The findings of this study support the validity of a model of approaches with four correlated factors: superficial approach, superficial intermediate approach, deep intermediate approach and deep approach. This model was tested by means of item confirmatory factor analysis showing acceptable fit, (CFI = .946, RMSEA = .037, 95% CI [.037, .042]) and average factor loads of .66 (superficial approach), .34 (superficial intermediate approach), .41 (deep intermediate approach) and .50 (deep approach). This study also showed that SLAT-Thinking produces reliable scores for superficial (.82), deep intermediate (.66) and deep (.69), according to Cronbach's alpha, and reliable scores for superficial approach (.61), according to McDonald's omega. Evidence of configural, metric and scalar invariance in two different samples was also supported in this study.

Despite the evidence of internal validity, SLAT-Thinking has not yet been studied for external validity. External validity is an important step, as it assesses whether or not the test is capable of producing empirical evidence that supports theoretically established relationships between the construct measured by the test and related variables. The theory

of learning approaches postulates that the deep approach provides more effective learning, while the superficial approach provides poor quality learning. Meta-analysis studies show that approaches are associated with academic performance (Richardson et al., 2012; Watkins, 2001). Watkins (2001) found mean correlations of .14 and -.18 between academic performance and the deep and superficial approaches, respectively. Richardson et al. (2012) found similar results, identifying mean correlations of .16 and -.11 between the deep and superficial approaches. Despite the small effect size, these studies show that there is a predictive relationship between approaches and academic performance. For this reason, this study examines whether SLAT-Thinking predicts academic performance.

Adding to the external validity analyzes, this paper will also study the incremental validity of SLAT-Thinking. Most predictors of academic performance assume the need for an active interaction between the subject and objects of knowledge (Cardoso et al., 2019; Pereira et al. 2019). This need is supported by constructivist theories (Golino, Gomes, Commons, et al., 2014; Gomes, 2007, 2010a; Gomes & Borges, 2009a; Pires & Gomes, 2018) and by neuropsychology (Dias et al., 2015; Reppold et al., 2015). Although the literature indicates that there are many predictors for academic performance, such as students' approaches to learning (Gomes, 2010c, 2011a, 2013; Gomes & Golino, 2012b; Gomes et al., 2011, 2022), metacognition (Gomes & Golino, 2014; Gomes et al. 2014), students' beliefs about the teaching-learning processes (Alves et al., 2012; Gomes & Borges, 2008a), motivation for learning (Gomes & Gjikuria, 2018), academic self-reference (Costa et al., 2017) and learning styles (Gomes & Marques, 2016; Gomes, Marques, et al., 2014), intelligence has a prominent place (Alves et al., 2016, 2017, 2018; Gomes, 2011b, 2012; Gomes & Borges, 2007, 2008b, 2009a, 2009b; Gomes & Golino, 2012a, 2015; Muniz et al., 2016; Valentini et al., 2015).

Intelligence is recognized as one of the variables which is traditionally studied in order to predict academic performance. This study examines whether SLAT-Thinking still increases the prediction of student performance in the presence of intelligence as a control variable. This study also examines whether SLAT-Thinking adds in predicting academic performance in the presence of a traditional measure of self-reporting in approaches as control variable. In summary, we expect that SLAT-Thinking will predict academic achievement and provide incremental validity over traditional self-report instruments and intelligence.

Method

Participants

This sample includes 655 students from private (n = 391, 59.7%) and public (n = 264, 40.3%) Brazilian institutions from the three broad areas of knowledge, biological sciences (n = 77, 11.8%), exact sciences (n = 290, 44.3%) and humanities (n = 288, 44.0%). The sample consists mainly of female participants (n = 345, 52.7%) with an average age of 23.7 (SD = 6.9) years, ranging from a minimum age of 17 to a maximum of 68 years.

The final multiple regression model consisted of two predictors, a sample size of 223, and an R² of 7.4%. We applied the Beta (Type II Error Rate) Calculator for Multiple Regression (Soper, 2023) to this model, using a type I error rate of 5%. We found a type II error rate of 2.8% for the estimated R², which is considerably lower than the standard reference criterion of 20% (Banerjee et al., 2009).

Data collection procedures

The data used in this study come from two independent sampling collections carried out in 2019. One collection included 513 students and the other included 142 students. The projects were approved by different ethics committees of Brazilian institutions (CAAE: 06353219.9.0000.5149 e CAAE: 73453317.1.0000.0118) and the collections complied with ethical principles. These collections involved the application of SLAT-Thinking, the Inductive Reasoning Development Test and the Learning Approach Scale. In order to have an indicator of academic performance, this collection also obtained the marks of the National Exam of Upper Secondary Education (ENEM) from 239 students.

Instruments

SLAT-Thinking

SLAT-Thinking was created in 2018 by Gomes and Linhares (2018) to assess approaches in identifying the author's thinking. The test is intended to evaluate students

with at least incomplete high school and measures four levels of approaches: superficial, superficial intermediate, deep intermediate and deep. The test consists of two texts of similar sizes and 12 items related to each text. Each item consists of a proposition that may or may not represent the author's thinking. The respondent must mark one of three response options: E (express the author's thought), N (not express the author's thought) and Z (there is no way to answer). There is only one correct alternative, and each hit is recorded as 1, and each error is recorded as 0.

Inductive Reasoning Development Test (Teste de Raciocínio Indutivo - TDRI)

TDRI emerged from the Battery of Higher-Order Cognitive Factors (Bateria de Fatores Cognitivos de Alta-Ordem - BAFACALO) (Golino & Gomes, 2014). BAFACALO includes 18 tests that measure both the general factor and six broad factors of the Cattel-Horn-Carroll (CHC) model. BAFACALO shows evidence of external and internal validity (Alves et al., 2012; Gomes, 2010a, 2010b, 2011b, 2012; Gomes & Borges, 2009a, 2009b, 2009c; Gomes, de Araújo, et al. 2014; Gomes & Golino, 2012a, 2012b, 2015). TDRI derives from the Inductive Reasoning Test, a BAFACALO test that measures inductive reasoning. The Inductive Reasoning Test items have groups of letters ordered according to a rule. The respondent's task is to discover the rule and point out the alternative that does not follow it. TDRI is structured similarly to the Inductive Reasoning Test. The difference is that, in addition to the CHC model, it takes the hierarchical complexity model as a reference. By combining a psychometric model of intelligence with a hierarchical complexity model, TDRI allows the identification of different stages of intelligence development (Golino, Gomes, Commons, et al., 2014).

TDRI measures seven stages of inductive reasoning: single representation, representational mapping, representational system, single abstraction, abstract mapping, abstract system and metasystematic. Each stage is assessed by using eight items. Each item has groups of letters ordered according to a rule. The respondent must discover the rule and indicate the alternative that does not follow it. There is only one correct alternative, and each hit is recorded as 1, while each error is recorded as 0. TDRI has evidence of validity and reliability in Brazilian samples (Golino & Gomes, 2019; Golino,

Gomes, et al., 2014; Gomes, Golino, et al., 2014). The model tested in this study had a sample of 488 students and is characterized by a general factor of inductive reasoning and four specific factors, all orthogonal to each other. The general factor explains the variance of the 56 items and each specific factor explains 8 items. These specific factors represent the first four stages of TDRI. Stages 5, 6 and 7 were not defined in the model due to the limitations of the sample; few people got the items in these stages right, making their identification unfeasible. The model presented an acceptable fit (CFI = .997, RMSEA = .060, 95% CI [.058, .062]), and average factor loads of .44 (general factor), .90 (stage 1), .78 (stage 2), .73 (stage 3) and .71 (stage 4). TDRI produced reliable scores for the general factor (.94), stage 1 (.99), stage 2 (.97), stage 3 (.97) and stage 4 (.98) according to Cronbach's alpha, and reliable scores for the general factor (.66), stage 1 (.64) and stage 2 (.60) according to McDonald's omega.

Learning Approach Scale (Escala de Abordagem de Aprendizagem - EABAP)

EABAP is a self-report questionnaire that measures learning approaches in people who have at least incomplete elementary education. The instrument consists of 17 items that represent motivations and strategies related to classroom and study. EABAP includes 8 items that measure superficial approach and 9 items that measure deep approach. The respondent must answer how much each behavior is present in his life, answering on a Likert-type scale from 1 to 5 where 1 represents "not at all" and 5 represents "totally". EABAP shows evidence of validity and reliability in Brazilian samples of primary and secondary education (Gomes, 2010c, 2011a, 2013; Gomes & Golino, 2012b; Gomes et al., 2011). The EABAP model with the two correlated factors had a sample of 648 students and showed an acceptable fit (CFI = .968 and RMSEA = .073, 95% CI [.066, .079]). The deep approach and the superficial approach correlated at -.60 and had mean factor loads of .64. The scores were found to be reliable, according to Cronbach's alpha, with .84 and .86 for superficial and deep approach and with .83 and .86 according to McDonald's omega.

Data Analysis

All analyzes were performed using version 4.0.2 of the R software (R Core Team, 2020), and involved three steps. The analyzes used the scores of SLAT-Thinking, EABAP and TDRI. The scores of superficial, deep intermediate and deep approaches of SLAT-Thinking, as well the EABAP scores were calculated from the relative average of responses to the items. For example, the superficial approach of SLAT-Thinking includes 6 items. If the respondent hits four of these items, then he will have a score of $4/6 = .67$. In turn, the superficial approach of EABAP includes eight items. If the respondent selects the value of 1 on the Likert-type scale for all these items, then their score will be $8/8 = 1$. TDRI scores consist of the sum of correct answers divided by eight (number of items per stage). For example, if the respondent hits 40 items, then his score is $40/8 = 5$. This score indicates the respondent's stage on inductive reasoning. The overall ENEM score can vary from 0 to 1000, so that the higher the score, the better the performance.

The first stage includes the descriptive statistics of the variables. In this step, the mean, the standard deviation, the minimum and maximum values, the kurtosis, the asymmetry and the correlation matrix are presented.

The second stage of the analysis examines by means of multiple linear regression whether or not SLAT-Thinking predicts academic performance. Versions 0.5.3 and 3.0.9 of the olsrr (Hebbali, 2020) and car (Fox & Weisberg, 2019) packages were used in this analysis. The tested model has the ENEM global score as a dependent variable, and inductive reasoning and the superficial and deep approaches measured by EABAP and superficial, deep intermediate and deep SLAT-Thinking approaches as independent variables. The stepwise forward method was used in order to select only the variables that increase the explanation of the dependent variable. The variance inflation factor (VIF) was inspected to examine multicollinearity. The Shapiro-Wilk test, kurtosis and skewness were used to assess the normality of the residues, just as the Bonferroni Outlier test was used to examine the presence of outliers and the score test was used to assess homoscedasticity.

In the third step of the analysis a regression tree is performed, using the rpart package (Therneau & Atkinson, 2019) and involving the same variables as the previous step. The tree regression model does not assume that the relationships between variables are necessarily linear, nor does it assume important data characteristics, such as normality of the dependent variable, homoscedasticity and independence of predictors. Although the literature suggests the use of pruning to minimize overfitting (Osei-Bryson, 2008), this procedure underestimates the correct number of leaves in small samples and overestimates in large samples. For this reason, this procedure was not used. For more details on the CART algorithm, see Gomes and Almeida (2017), Gomes and Jelihovschi (2019, 2020), Gomes, Amantes, et al. (2020), as Gomes, Lemos, et al. (2020).

Results and discussion

The results of the descriptive analyzes can be seen in Table 1. The mean of the deep self-reported approach is higher than the mean of the superficial self-reported approach, since the 95% confidence interval of the deep approach ($3.82 \pm 0.03 * 1.96 = 3.76, 3.88$) does not overlaps with the confidence interval of the self-reported superficial approach ($2.62 \pm 0.03 * 1.96 = 2.56, 2.68$). This indicates that the participants in this study perceive themselves as deeper than superficial. In addition, as their average score is 3.82, and this value is very close to point 4 of the scale, it can be said that the participants have a perception that deep behaviors are frequent in their repertoire. In turn, the participants believe that the superficial approach behaviors are moderately present in their repertoire. The average of the performance in superficial approach is greater ($.80 \pm .01 * 1.96 = .78, .82$) than the averages of the deep intermediate ($.32 \pm .01 * 1.96 = .30, .34$) and deep ($.11 \pm .01 * 1.96 = .09, .13$). This indicates that the participants in this study predominantly achieve the superficial approach to the ability assessed by SLAT-Thinking. The TDRI average indicates that the sample participants are close to the fourth stage of inductive reasoning, the singular abstract stage. Considering that the ENEM scale is prepared by the National Institute of Educational Studies and Research Anísio Teixeira (INEP) to have an average of 500 and a standard deviation of 100 (BRASIL/INEP, 2011), it appears that the participants in this research had a performance above one standard

deviation in relation to the scale average, indicating a performance above the national average.

Statistically significant correlations (p < .05) are highlighted in bold.

Table 1

Descriptive Statistics of the model's variables

Variable	ENEM	D.A. EABAP	S.A. EABAP	S.A. SLAT	I.D.A. SLAT	D.A. SLAT	TDRI
ENEM	1.00						
D.A. EABAP	.05 [-.07, .16]	1.00					
S.A. EABAP	-.02 [-.14, .12]	.00 [-.17, .15]	1.00				
S.A. SLAT	**.25** **[.15, .38]**	**.10** **[.02, .20]**	.00 [-.06, .09]	1.00			
I.D.A. SLAT	.01 [-.12, .11]	.04 [-.04, .11]	.01 [-.05, .09]	**.24** **[.18, .30]**	1.00		
D.A. SLAT	-.06 [-.18, .08]	.03 [-.05, .10]	-.01 [-.08, .08]	-.08 [-.17, .00]	**.56** **[.51, .61]**	1.00	
TDRI	**.16** **[.03, .28]**	.05 [-.02, .13]	**-.20** **[-.25, -.11]**	**.14** **[.06, .19]**	.00 [-.07, .08]	**-.10** **[-.17, -.02]**	1.00

	N	mean (SD)	SE	min.	max.	kurtosis	skewness
ENEM	239	616.58 (105.56)	6.83	300	832	-0.15	-0.42
D.A. EABAP	655	3.82 (0.86)	0.03	0	5	6.94	-2.11
S.A. EABAP	655	2.62 (0.88)	0.03	0	5	0.66	-0.19
S.A. SLAT	655	0.80 (0.23)	0.01	0	1	1.66	-1.39
I.D.A. SLAT	655	0.32 (0.23)	0.01	0	1	-0.05	0.67
D.A. SLAT	655	0.11 (0.16)	0.01	0	1	3.48	1.67
TDRI	488	3.78 (0.96)	0.04	1	6	-0.45	-0.12

Note. D.A. EABAP = deep approach from EABAP, S.A EABAP = superficial approach from EABAP, S.A. SLAT = superficial approach from Slat-Thinking, I.D.A. SLAT = intermediate deep approach from SLAT-Thinking, D.A. SLAT = deep approach from SLAT-Thinking, SD = standard deviation, SE = standard error, min = minimum, max = maximum.

The ENEM global score correlates positively and weakly with the superficial approach of SLAT-Thinking and with inductive reasoning (TDRI). Among the dimensions of SLAT-Thinking and EABAP there is only a statistically significant correlation between the deep approach of EABAP and the superficial approach of SLAT-Thinking (r = .10). This indicates that the dimensions of the two instruments are orthogonal to each other. This orthogonality is not a problem, as the EABAP evaluates approaches in a broad context that involves the diversity of study and learning behaviors in the classroom. In turn, SLAT-Thinking assesses the individual's approaches to identify the author's thinking in a given text. Since the superficial approach of SLAT-Thinking correlates positively with the overall score of ENEM (r = .25) and with inductive reasoning (r = .14), and as we contrast this result with the negative correlation between the superficial approach of EABAP and inductive reasoning (r = -.20), it can be seen that the superficial approach derived from performance is different from the superficial self-reported approach. If the motivation in the self-reported superficial approach involves low engagement, this does not necessarily occur in the superficial approach measured by performance. Unlike the meta-analyzes presented in the introduction to this article, no statistically significant correlations were found between self-reported approaches and academic performance. On the other hand, the size of the correlation between academic performance and the superficial approach of SLAT-Thinking is similar to those found in meta-analyzes.

The step forward multiple regression analysis indicated that only the superficial approach of SLAT-Thinking and inductive reasoning explain the ENEM global score. The VIF was 1 for the two independent variables, indicating that the model does not have multicollinearity. This was already expected, given that the scores of the inductive reasoning and the superficial approach of SLAT-Thinking are weakly correlated (Table 1). The residues showed normal distribution (Shapiro-Wilk test, W = .988, p = .069, kurtosis = 0.23, skewness = -0.34), as well as constant variance (non-constant variance, $\chi^2 (1) = 1.00$, p = .316). The model has no outliers (Bonferroni outlier test, rstudent = -2,943 and Bonferroni p = .803).

The model's intercept was 441.67, CI 95% [364.70, 518.64], indicating that if the participant misses all items in the superficial approach of SLAT-Thinking and TDRI, he will get an overall score in the ENEM of 441.67 points. The slope of 123.80, CI 95% [56.96, 190.65] for the superficial approach of SLAT-Thinking indicates that by getting all the items in that domain right, it would add 123.80 points to the overall ENEM score, producing a score of 565.47 points. The slope of 16.76, CI 95% [2.946, 30.56] for inductive reasoning indicates that at each dominated stage of inductive reasoning there is an increase of 16.76 points in the overall score of ENEM.

The superficial approach of SLAT-Thinking explained .0592 (adjusted R^2 = 5.92%) of the variance of the overall ENEM score, while inductive reasoning increased the explanation of variance by .015 (adjusted R^2 = 1.5%), so that the model explained .0742 (adjusted R^2 = 7.42%) of the variance of the overall ENEM score. If we take the Cohen (1988) criterion as a reference, and knowing that the R^2 of 7.42% is equal to r of .272, we conclude that the variance explained by the model has a weak to moderate size. These results indicate that SLAT-Thinking shows evidence of predictive and incremental validity, insofar as the superficial approach is the main predictor of the ENEM global score.

The tree produced by the CART algorithm is shown in Figure 1. Each oval element in Figure 1 is a terminal node, in other words, a leaf of the tree. Within these elements there is a number that represents the predicted score of the people who are allocated by the CART algorithm to belong to that leaf. For example, the upper left extreme leaf shows the number 510, indicating that the participants who are part of that leaf have, according to the predictive model, a score of 510 points on the overall ENEM score. In turn, the number at the bottom of the oval element indicates the relative percentage of the sample. In our example, the number 7% indicates that 7% of the participants are part of this leaf.

The tree must be read from top to bottom. Let us take again the example of the upper left extreme sheet. To understand the characteristics of this group of people, it is necessary to observe the initial group at the top of Figure 1 and follow the lines that lead to the sheet. The upper node is composed of all persons in the sample. This node was broken into two parts using the .92 score of the superficial approach of SLAT-Thinking.

People who scored less than .92, were allocated to the leaves on the left of Figure 1, while people who scored at least .92 were allocated to the leaves on the right. Note that the upper left extreme sheet is made up of people who scored less than .92 in a superficial approach to SLAT-Thinking. However, this sheet on the extreme left is not only characterized by the score in the superficial approach of SLAT-Thinking. Only people with a score below 2.7 points in inductive reasoning are part of this group. In short, this group is characterized by people who do not master the superficial approach and have not reached the stage of a system of representation of inductive reasoning.

CART produced 19 splits, 20 leaves and a relative error of .65, indicating that the model explained 35% of the variance of the ENEM global score. We simulated 100 regression trees with pruning for three different sample sizes, the first with 239 observations (sample size for this analysis), the second with 10,000 observations and the third with 100,000 observations. The simulations considered as true the means and standard deviations of the variables of the tested model, as well as the cutoff points of the tree nodes. The averages of the number of leaves produced by the 100 simulations for each sample size were calculated.

The simulation with 239 observations produced pruned trees with an average of 12.21 (SD = 3.86) leaves, ranging from 2 to 20 leaves. By taking the original tree as the true tree, this result indicates an evident underestimation of the true number of leaves. The simulation with 10,000 observations produced pruned trees with an average of 53.62 (SD = 12.11) leaves, ranging from 36 to 85. The simulation with 100,000 observations produced pruned trees with an average of 66.5 (SD = 12.11) leaves, ranging from 62 to 71. These results indicate that, for large samples, the pruning procedure overestimates the correct number of leaves relative to the original tree. In short, the simulations indicate that the pruning does not estimate the correct number of leaves and does not eliminate the problem of overfitting. We believe that the best way to assess the consistency of the leaves identified by the tree in this study is to investigate whether the model tested in this study replicates in other samples.

Figure 1

Regression tree
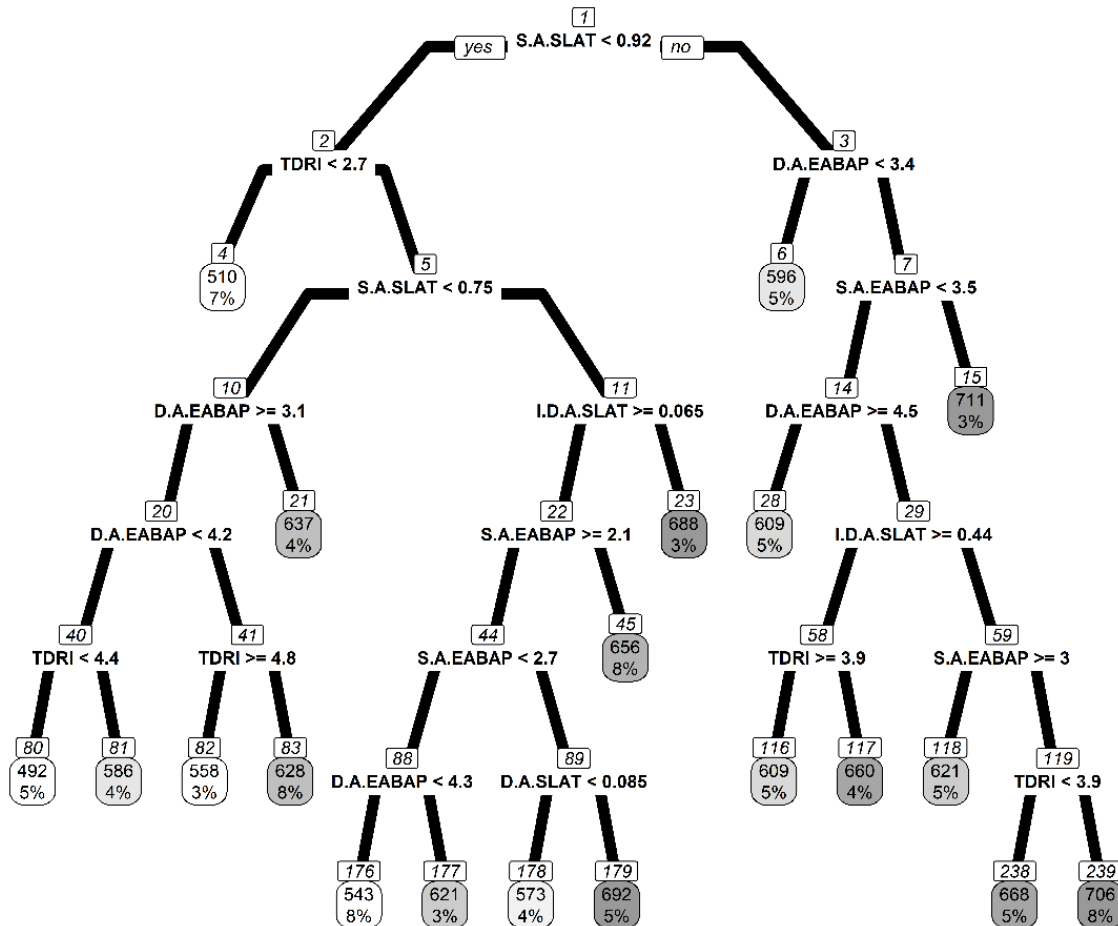
Note. D.A. EABAP = deep approach from EABAP, S.A EABAP = superficial approach from S.A EABAP, S.A. SLAT = superficial approach from Slat-Thinking, I.D.A. SLAT = intermediate deep approach from SLAT-Thinking, D.A. SLAT = deep approach from SLAT-Thinking.

The regression tree (Figure 1) indicates that students with the lowest overall ENEM score (492) have a score slightly below the national average. These students: (1) do not master the superficial approach of SLAT-Thinking (S.A. SLAT < .75), (2) believe that behaviors of deep approach in the context of the classroom and study are moderate to very present in their repertoire (D.A. EABAP > = 3.1 and D.A. EABAP < 4.2) and (3) and present, at most, the singular abstraction stage in inductive reasoning (TDRI < 4.4). In turn, students with the highest ENEM global score (711) believe that they have, at least, moderate behaviors of: (1) deep approach (D.A. EABAP > 3.4) and (2) superficial approach (S.A. EABAP > 3.5) in the context of the classroom and study. In short, the group with the highest ENEM score is the group that reports combining the two approaches, maximizing performance. This combination is evidence favorable to the

strategic approach, which is in line with Biggs' (1985) argument that the student highly motivated to get good grades combines strategies from both approaches to achieve better performance.

Since it is the first variable chosen by the CART algorithm to break the data into nodes, the superficial approach of SLAT-Thinking is the most important predictor variable in the model. A greater superficial approach implies a higher overall score for ENEM.

The tree shows that having a greater deep intermediate approach to SLAT-Thinking is associated with lower overall score of ENEM. This result is not what we would expect, since, in theory, the deep intermediate approach should be positively associated with academic performance. It is important to note whether this result is replicated in new studies, or it is just a peculiarity or mere randomness of the sample in this study.

The deep approach of SLAT-Thinking is positively associated with the overall score of ENEM, conditioned to the superficial approach of SLAT-Thinking, to the inductive reasoning, to the deep intermediate approach and to the superficial approach of EABAP. Getting at least 8.5% of the items in this approach results in a score from 573 to 692 points in the ENEM, a difference of 119 points, indicating an increase of more than one standard deviation.

Inductive reasoning has non-linear relationships with the overall score of ENEM. The stages of the representation map (stage 2), representation system (stage 3), singular abstraction (stage 4) and abstraction map (stage 5) differentiate the overall score of ENEM.

The deep approach of EABAP has non-linear relationships with the overall score of ENEM. There are situations in which reporting a more profound approach in the context of the classroom and study implies a higher ENEM score, while there are situations in which this relationship is reversed. The same is true for the superficial approach of EABAP.

Conclusion

This article examined the predictive and incremental validity of SLAT-Thinking relative to the overall score of ENEM. The superficial approach of SLAT-Thinking was the main predictor and inductive reasoning the second most important predictor, both in the linear multiple regression model and in the tree regression model.

While the multiple linear regression model explained 7.42% of the variance of the overall ENEM score, the tree regression model explained 35% of the outcome variance. In the case that the regression tree does not have too much overfitting, which is only possible to know by replicating this study in several other samples, it can be said that the tree regression variance explanation is much higher than the multiple linear regression's due to the predictors having pronounced non-linear relationships as an outcome. The results of the regression tree indicated that the superficial and deep approaches of EABAP and inductive reasoning have non-linear relationships with the ENEM scores.

This study has a shortcoming which is the lack heterogeneity in the sample relative to the performance in SLAT-Thinking approaches and in inductive reasoning. Few participants in the sample of this study hit the correct answers about the items of the deep intermediate, deep approaches and the items of the more advanced stages of TDRI. Only 98 participants hit more than half of the items of deep intermediate approach, while only 8 participants hit more than half of the items of deep approach of SLAT-Thinking. Only 61 people reached the abstraction map stage in inductive reasoning, only one reached the abstraction system stage and none reached the meta-systematic stage.

It is important to note that TDRI has 5 multiple-choice options, while SLAT-Thinking has only 3 response options, one of which is very unlikely to be answered because it is an option that the respondent claims to have no way to answer. This option is not part of the answer key and has no logical support, and may be selected only by people with a very high superficial approach. Therefore, the respondent has approximately a 50% chance of hitting a SLAT-Thinking item by merely choosing the answer at random and it is very likely that a great number of participants with high performance in the intermediate-deep approach and in the deep approach have had a high performance at random. This feature of the test shall be changed in future versions. A new version of SLAT-Thinking with many multiple-choice options may reverse this strong chance of hitting at random, making the assessment of deep approach levels more reliable. This would also allow us to analyze whether the lack of correlation between the deep approach and the ENEM scores is related to the characteristic of SLAT-Thinking itself.

Despite the aforementioned shortcomings, this study brings evidence of the predictive and incremental validity of SLAT-Thinking. This test increases the prediction of the overall score of ENEM, taking as control both the inductive reasoning and the self-report of approaches.

The results have shown that SLAT-Thinking predicts student performance better than inductive reasoning, they also show that the self-report of the learning approach had no predictive role in the regression linear model. This clearly shows that the performance test is superior to the test based on self-report to predict student performance. The results also show that the higher the performance in the superficial approach of SLAT-Thinking, the greater the performance in ENEM. This appears to look inadequate, considering the evidence from the theory of learning approaches which point out that the superficial approach has a negative correlation with academic performance. Nonetheless, it must be taken into account that all evidence obtained so far by the area comes from self-report instruments. The negative correlation makes sense when we think about the superficial approach items contained in the self-report instruments. We must remember that SLAT-Thinking is the first test of learning approaches based on performance. Items of superficial approach, based on the ability to identify the author's thinking in a given text, are apparently easy items, in which a faster and superficial reading allows them to be hit by the respondents. However, missing these items is not a positive thing in theory. It is interesting to note that performance tests change the logic and perspective of expecting that the superficial approach is necessarily opposed to deep approach and performance.

It makes no theoretical sense that a respondent misses the easy items that require only a quick and superficial reading and hits the difficult items of the deep approach, which demand forming non-obvious relationships and reading details in the arguments. In this sense, this article brings a new conceptual perspective to the field of studies in learning approaches.

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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://revistaepsi.com/artigo/2012-ano2-volume1-artigo5/

Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal*, *18*(2), 127-131. https://doi.org/10.4103/0972-6748.62274

Biggs, J. B. (1985). The role of meta-learning in study process. *British Journal of Educational Psychology*, *55*, 185-212. https://doi.org/10.1111/j.2044-8279.1985.tb02625.x

BRASIL/INEP (2011). *Nota Técnica Procedimento de cálculo das notas do Enem.* INEP, Brasília. http://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_procedimento_de_calculo_das_notas_enem_2.pdf

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.), Erlbaum.

Costa, B. C. G., Gomes, C. M. A., & Fleith, D. S. (2017). Validade da Escala de Cognições Acadêmicas Autorreferentes: autoconceito, autoeficácia, autoestima e valor. *Avaliação Psicológica*, *16*(1), 87-97. https://doi.org/10.15689/ap.2017.1601.10

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. https://doi.org/10.15348/1980-6906/psicologia.v17n2p140-152

Fox, J., & Weisberg, S. (2019). *car: Companion to Applied Regression, R package (version 3.0-9)* [Computer software]. https://cran.r-project.org/web/packages/car/car.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

Golino, H. F., & Gomes, C. M. A. (2019). *TDRI: Teste de Desenvolvimento do Raciocínio Indutivo,* 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. https://doi.org/10.4236/psych.2014.518207

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. https://doi.org/10.1037/h0100589

Gomes, C. M. A. (2007). Softwares educacionais podem ser instrumentos psicológicos. *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. *Psicologia em Estudo*, *15*(4), 841-849. 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. 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. 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. 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). *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). *Revista Interinstitucional de Psicologia*, *5*(2), 294-316. 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 (PUCRS. Online)*, *44*(2), 193-203. http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/11371

Gomes, C. M. A., & Almeida, L. S. (2017). Advocating the broad use of the decision tree method in education. *Practical Assessment, Research & Evaluation, 22*(10), 1-10. https://pareonline.net/getvn.asp?v=22&n=10

Gomes, C. M. A., Amantes, A., & Jelihovschi, E. G. (2020). Applying the Regression Tree Method to Predict Students' Science Achievement. *Trends in Psychology*. *28*(1), 99-177. https://doi.org/10.9788/s43076-019-00002-5

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. 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. http://www.cienciasecognicao.org/revista/index.php/cec/article/view/60

Gomes, C. M. A., & Borges, O. (2008b). Qualidades psicométricas de um conjunto de 45 testes cognitivos. *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. *Estudos em Avaliação Educacional, 20*(42), 73-88. https://doi.org/10.18222/eae204220092060

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

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

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. https://doi.org/10.1037/h0101078

Gomes, C. M. A., Farias, H. B., & Jelihovschi, E. G. (2022). Approaches to learning does matter to predict academic achievement. *Revista de Psicología*, *40*(2), 905-933. https://doi.org/10.18800/psico.202202.010

Gomes, C. M. A., & Gjikuria, E. (2018). Structural Validity of the School Aspirations Questionnaire (SAQ). *Psicologia: Teoria e Pesquisa*, *34*, e3438. 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? *Psicologia: teoria e prática*, *14*(1), 126-139. 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. https://doi.org/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. https://doi.org/10.1590/1678-7153.201427307

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

Gomes, C. M. A., Golino, H. F., & Menezes, I. G. (2014). Predicting School Achievement Rather than Intelligence: Does Metacognition Matter? *Psychology, 5*, 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. *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., (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.

http://www.sciencedomain.org/reviewhistory.php?iid=488&id=21&aid=4724

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

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. https://doi.org/10.15689/ap.2020.1901.17737.10

Gomes, C. M. A. & Linhares, I. (2018). Investigação da validade de conteúdo do TAP-Pensamento. [Pôster]. *I Encontro Anual da Rede Nacional de Ciência para Educação (CPE)*. https://doi.org/10.13140/RG.2.2.31110.40006

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. https://doi.org/10.15689/ap.2016.1503.05

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. https://revistaepsi.com/artigo/2013-2014-ano3-volume2-artigo3/

Gomes, C. M. A., Quadros, J. S., Araujo, J., & Jelihovschi, E. G. (2020). Measuring students' learning approaches through achievement: structural validity of SLAT-

Thinking. *Estudos de Psicologia (Natal)*, *25*(1), 33-43. https://doi.org/10.22491/1678-4669.20200004

Hebbali, A. (2020). *olsrr: Tools for Building OLS Regression Models, R package (version 0.5.3)* [Computer software]. https://cran.r-project.org/web/packages/olsrr/olsrr.pdf

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

Osei-Bryson, K. (2008). Post-pruning in regression tree induction: An integrated approach. *Expert Systems with Applications, 34*, 1481-1490. https://doi.org/10.1016/j.eswa.2007.01.017

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. https://doi.org/10.5281/zenodo.3477577

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

R Core Team. (2020). *R: A language and environment for statistical computing (version 4.0)* [Computer software]. https://www.R-project.org/

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353-387. https://doi.org/10.1037/a0026838

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. https://doi.org/10.15348/1980-6906/psicologia.v17n2p94-106

Soper, D.S. (2023). *Beta (Type II Error Rate) Calculator for Multiple Regression* [Software]. https://www.danielsoper.com/statcalc

Therneau. T., & Atkinson, B. (2019). *rpart: Recursive Partitioning and Regression Trees R package (version 4.1-15)* [Computer software]. https://cran.r-project.org/web/packages/rpart/rpart.pdf

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. https://doi.org/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), Lawrence Erlbaum Associates.