A new methodology to evaluate factor scores: internal and external correlational accuracy

Heitor Blesa Farias
Universidade Federal de Minas Gerais, Brazil
ORCID: https://orcid.org/0000-0002-8090-4012
E-mail: heitorblesa@gmail.com

Cristiano Mauro Assis Gomes
Universidade Federal de Minas Gerais, Brazil,
ORCID: https://orcid.org/0000-0003-3939-5807
E-mail: cristianomaurogomes@gmail.com

Enio Galinkin Jelihovschi,
Universidade Estadual de Santa Cruz, Brazil,
ORCID: https://orcid.org/0000-0002-7286-1198
E-mail: eniojelihovs@gmail.com

Corresponding Author: Heitor Blesa Farias.
Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
Acknowledgements: Cristiano Mauro Assis Gomes: Productivity Fellowship, CNPq, Brazil.

Abstract: In factor analysis, the indeterminacy of factor scores brings the possibility to produce multiple solutions, which often do not reproduce the true correlations of the factors in a measurement model. Grice (2001) emphasizes the need to evaluate the similarity between the correlations of the factors in the measurement model and those of their factor scores, terming this similarity correlational accuracy. Existing factor score techniques address this issue within a single measurement model, posing a limitation when multiple models are relevant. Moreover, Grice's proposal lacks a well-defined methodological framework. This article addresses these limitations by introducing two systematic categories of analysis: internal and external correlational accuracy. In the first of these, we create a well-defined methodological path for Grice's proposal. In the second, we create a way of evaluating factor scores in the context of various measurement models. A step-by-step method and examples are presented.

Keywords: Correlational Accuracy; Factor score indeterminacy; Factor analysis; Psychometrics; Tests.

1. Introduction

Studies of factor score indeterminacy show that the factorial scores of measurement models tend to be biased, as they usually produce multiple solutions (Ferrando & Lorenzo-Seva, 2018), which mostly do not reproduce the true correlations between the factors (Croon, 2002; Devlieger et al., 2016; Grice, 2001; Skrondal & Laake, 2001; Steiger, 1979). This is quite problematic since the use of factor scores has become a worldwide trend of broad practice in predictive studies (Devlieger et al., 2019).

Grice (2001) states that this is unacceptable and defends the need for an assessment that examines the similarity between the correlations of the factors in the measurement model and the correlations of their factor scores, calling this similarity correlational accuracy. In mathematical terms, Grice's (2001) correlational accuracy indicates the extent to which the correlations among the estimated factor scores match the correlations among the factors themselves. This similarity is measured by the elementwise difference between the factor score correlation matrix and the true factor correlation matrix. This difference indicates the degree of bias in the factorial scores.

Factor score bias has important implications in research and clinical practice. Suppose a clinician applies a test measuring perfectionism and anxiety to her patients. She wants to assess whether perfectionism plays an important predictive role in explaining her patients' anxiety. She performs a factor analysis with two factors (perfectionism and anxiety) and calculates the factor scores. If the true correlation between perfectionism and anxiety is .80, then perfectionism predicts 64% of the variance in anxiety. On the other hand, if the factor scores for perfectionism and anxiety show a correlation of .50, there is a bias of -.30 (Δ = .80 - .50 = -.30). This bias will lead the clinician to wrongly conclude that perfectionism only predicts 25% of patients' anxiety, losing 39% of the true prediction.

Grice's (2001) warning and the factor score techniques created to solve this problem (i.e. Beauducel et al., 2023; McDonald, 1981; Ten Berge et al., 1999) were designed for the context of a single measurement model. This limitation of context creates an important problem, as there are frequent situations in which it is not appropriate to run a single measurement model. For example, there is a lot of evidence in the psychometric literature that fit indices are not good for detecting local fit in confirmatory factor analysis and structural equation modeling (Thoemmes et al., 2018). If a model has a factor structure with three tests (A, B and C), fit indices are not able to properly assess whether the factor structure of test A, B or C has adequate fit. To obtain fit indices that properly assess the factor structure of each test, the researcher needs to break down the complete model into a model for each test. Another frequent situation in which it is not appropriate to run a single measurement model occurs when the sample is not large enough to properly estimate the parameters of a complex factorial structure (Jobst et al., 2023). In addition to the limitation of the context, Grice's (2001) proposal does not define a well-defined methodological path.

In this article, we created two systematic categories of analysis: internal and external correlational accuracy. First, we created a well-defined methodological path for Grice's (2001) proposal. Then, we developed a systematic way of evaluating the factor scores in the context of various measurement models. We also present a step-by-step method and examples of its application.

2. Methodology

Let's assume a multi-factor model, with all the factors being estimated in a single measurement model, at the same time. If the factor scores faithfully reproduce the true latent correlations between these factors, we will have perfect internal correlational accuracy. In this case, we call internal correlational accuracy the assessment of the degree to which the true latent correlations are reproduced in a single measurement model. For this category of systematic analysis, we must follow the following assumptions: (1) there must be a single measurement model, which can contain either the factor structure of part of a test, a whole test, or two or more tests; (2) factor scores must be extracted from this model; (3) true correlations between the factors are estimated via confirmatory factor analysis and structural equations modeling; (4) the measurement model must be multidimensional.

The well-defined methodological path for Grice's (2001) proposal is presented below:

1 The measurement model must be tested via confirmatory factor analysis or structural equations modeling.

2 The tested measurement model must have an acceptable fit. If not, test a new model. You can use your preferred fit indexes; we suggest using CFI $\geq .90$ and RMSEA $< .10$.

3 Having a model with acceptable fit, check the correlations between the factors, assumed to be the true ones.

4 Estimate the factor scores of the model, selecting one of the available factor generation techniques.

5 Calculate the correlations of the estimated factorial scores.

6 Calculate the distance of these correlations from the true correlations (Δ = factorial score correlation - true correlation) to estimate the correlational accuracy bias.

We call external correlational accuracy the assessment of the degree to which the true latent correlations are reproduced in the context of multiple measurement models. For this category of systematic analysis, we must follow the following assumptions: (1) there must be the analysis of three or more measurement models separately. Each separate model may be part of a factorial structure of a test, the complete factorial structure of a test, or it may contain the factorial structure of two or more tests; (2) True correlations among factors from separate models can be estimated using either pairwise confirmatory factor analysis or pairwise structural equations modeling. If models A, B, C, and D each have a single factor, we perform confirmatory factor analyses for pairs A-B, A-C, A-D, B-C, B-D, and C-D. The resulting estimated correlations between factors are considered accurate, creating a true correlation matrix.

The methodological path of external correlational accuracy is presented below:

1 Each separate measurement model must be tested via confirmatory factor analysis or structural equations modeling.

2 Each separate model must have acceptable fit. If not, a new separate model needs to be tested.

3 Confirmatory factor analyses or structural equations modeling of the pairwise models should be performed. When creating pairwise models, the factors in one model are correlated with those in another model.

4 Inspect the correlations between the factors in each pairwise model, assuming them to be true.

5 Estimate the factor scores of the separate models from step 2, selecting one of the available factor generation techniques.

6 Calculate the correlations of the estimated factor scores.

7 Calculate the distance of these correlations from the true correlations (Δ = factorial score correlation - true correlation).

3. Internal correlational accuracy: Example

We apply internal correlational accuracy to evaluate the bias of factorial scores from a Fluid Intelligence Kit (CTIF) measurement model. The CTIF consists of three tests, each of them measuring a specific reasoning ability: general reasoning, inductive reasoning and logical reasoning. In addition to specific abilities, the CTIF measures the broad ability of fluid intelligence (Details about the CTIF can be seen in Table 3).

Step 1 - The measurement model should be tested.

The CTIF measurement model tested is a bifactor model with the presence of one general latent variable, fluid intelligence, and three specific latent variables, inductive reasoning, logical reasoning, and general reasoning, all orthogonalized to each other. We applied item confirmatory factor analysis for this model with the Weighted Least Square Mean and Variance Adjusted (WLSMV) estimator and using the lavaan package (Rosseel, 2012).

Step 2 - The tested measurement model needs to have acceptable fit.

The tested model had acceptable fit (χ^2 [1506] = 3427.38, CFI = .926, RMSEA = .040 [.038 - .042]).

Step 3 - Inspect the correlations between the factors, assuming them to be true.

The model tested shows that all factors have zero true correlation.

Step 4 - Estimate the factor scores of the model, selecting one of the available factor generation techniques.

We used the lavaan package regression technique (Rosseel, 2012) to estimate the factor scores. We chose this technique because it is the default of the package and, consequently, the most widely used.

Step 5 and 6 - Calculate the correlations of the factor scores and the distance of these correlations from the true correlations to estimate the correlational accuracy bias.

Table 1 presents the calculations of steps 5 and 6 and reports the bias of the factorial scores of the CTIF measurement model. For example, the difference of the correlation of the factor scores of general reasoning (GR) and inductive reasoning (IR) from the true correlation is -.120 (see Table 1). The difference is obtained as follows: factorial score correlation [-.120] - true correlation [0] = -.120. When the value of the difference (Δ) is negative, the correlation of the factor scores is lower than the true correlation. When the difference (Δ) is positive, the correlation of the factorial scores is greater than the true correlation.

Table 1
Internal correlational accuracy bias of CTIF´s factor scores

Model		Factor scores correlations				Difference of the correlations of the factor scores from the true correlations			
		IR	**LR**	**GR**	**Gf**	**Δ IR**	**Δ LR**	**Δ GR**	**Δ Gf**
CTIF	**IR**	1				0			
	LR	.049	1			.049	0		
	GR	-.120	.156	1		-.120	.156	0	
	Gf	.120	.078	.159	1	.120	.078	.159	0

Note. IR = Inductive reasoning, LR = Logical reasoning, GR = General reasoning, Gf = Fluid intelligence.

4. External correlational accuracy: Example

We apply external correlational accuracy to evaluate the bias of factorial scores of the CTIF, the Approaches to Learning Scale (EABAP), and the CTCAM-Monitoring. The EABAP is a test designed to assess students' learning approaches, specifically measuring the deep and surface approaches. The CTCAM-Monitoring comprises items from the Academic Knowledge and Metacognition Testbooks. It evaluates the metacognitive ability of monitoring, which is the ability to detect errors while performing an activity. (See Table 3 for more information about the tests).

Step 1 - Each separate measurement model should be tested.

The CTIF measurement model was the same as the one used in the analysis of internal correlational accuracy. For the EABAP, a correlated factor measurement model was tested in which the deep and surface approach latent variables correlate. For CTCAM-Monitoring, a unidimensional measurement model was tested in which the latent variable is monitoring ability. We applied item confirmatory factor analysis with the Weighted Least Square Mean and Variance Adjusted (WLSMV) estimator, using the lavaan package (Rosseel, 2012).

Step 2 - Each separate model must have acceptable fit.

The three models showed acceptable fit, CTIF (χ^2 [1506] = 3427.38, CFI = .926, RMSEA = .040 [.038 - .042]), EABAP (χ^2 [118] = 622.22, CFI = .956, RMSEA = .073 [.068 - .079]) and CTCAM-Monitoring (χ^2 [2] = 1.304, CFI = 1.000, RMSEA = .000 [.000 - .062]).

Step 3 - Confirmatory factor analyses or structural equation modeling from the pairwise models.

Given that three measurement models are used in this example, there are three pairwise models developed:

1. CTIF (bifactor model) and EABAP (correlated factor model);
2. CTIF (bifactor model) and CTCAM-Monitoring (one-dimensional model);
3. EABAP (correlated factors model) and CTCAM-Monitoring (one-dimensional model).

We applied confirmatory factor analysis of items for each pairwise model with the Weighted Least Square Mean and Variance Adjusted (WLSMV) estimator, using the lavaan package

(Rosseel, 2012). For each pairwise model, the factors in one model are correlated with those in another model.

Step 4 - Inspect the correlations between the factors in each pairwise model, assuming them to be true.

Table 2 shows true factor correlations extracted from each pairwise model.

Step 5 - Calculate the factor scores for each model from step 2 using one of the available factor generation techniques.

We used the lavaan package regression technique (Rosseel, 2012) to estimate the factor scores for each model from step 2.

Step 6 and 7 - Calculate the correlations of the factorial scores and the distance of these correlations from the true correlations to estimate the correlational accuracy bias.

Table 2 shows the bias of the factorial scores of the CTIF, EABAP and CTCAM-Monitoring measurement models. Some biases were substantial. For example, the true correlation of deep approach (DA) with general reasoning (GR) is .34, but the correlation of the factorial scores was .18, producing a bias of Δ = -.16. The bias between fluid intelligence and monitoring was even greater (Δ = -.21).

Table 2

External correlational accuracy bias of CTIF, EABAP and Monitoring

Model		IR	LR	GR	Gf	DA	SA	Mon
Factor scores correlations	**IR**	1	.05	-.12	.12	-.01	.04	.07
	LR		1	.16	.08	.04	-.04	.16
	GR			1	.16	.18	-.20	.18
	Gf				1	.18	-.21	.56
	DA					1	-.76	.23
	SA						1	-.24
	Mon							1
Pairwise true correlations from pair models	**IR**	1	.00	.00	.00	.01	.05	.04
	LR		1	.00	.00	.06	-.05	.22
	GR			1	.00	.34	-.36	.19
	Gf				1	.15	-.22	.77
	DA					1	-.65	.31
	SA						1	-.35
	Mon							1
Difference between correlation matrices (Δ = Factor scores correlations – Pairwise true correlations from pair models)	**IR**	0	.05	-.12	.12	-.02	-.01	.03
	LR		0	.16	.08	-.02	.02	-.06
	GR			0	.16	-.16	.16	-.01
	Gf				0	.03	.02	-.21
	DA					0	-.11	-.08
	SA						0	.11
	Mon							0

Note. IR = Inductive reasoning, LR = Logical reasoning, GR = General reasoning, Gf = Fluid intelligence, DA = Deep Approach, SA = Surface Approach, Mon = Monitoring.

5. Pondering over a Cut-Off Point for Factor Score Bias

The correlational accuracy bias need not be zero. For example, a $\Delta = \pm.01$ does not represent relevant biases. We should think about the magnitude of the bias and the size of inadmissibility. A $\Delta = \pm.10$ seems impressive to us. Suppose factor A is used to predict factor B. If the true correlation between these factors is .54, then the proportion of the variance of B which is explained by A is 29.16%. If the factor score correlation between those factors were .44 (Δ = -.10), then the proportion of the variance of B which is explained by A would be 19.36%, losing 33.61% of the true prediction. The same is true if the factor score correlation between factor A and B were .64; in that case, the proportion of the variance of B explained by A would be 40.96%, overestimating the true prediction by 40.47%. It seems to us that a difference of up to .050 is acceptable. The user may also prefer not to use any cutoff point and just report the impact of the bias.

Table 3

The Data and Instruments of Our Examples

Sample	The sample is composed of 792 high school students (51.25% female and 55.55% enrolled in private schools); Five schools from Belo Horizonte and Viçosa, Minas Gerais, Brazil; Age ranged between 14 and 21 years-old (M = 16.3, SD = 1.00); Distributed homogeneously in high school grades (35.60% in first-year, 29.04% second-year and 34.36% third-year).
Fluid Intelligence Kit (CTIF)	CTIF is composed by Induction Test, Logical Reasoning Test, and General Reasoning Test (Gomes & Borges, 2009c); CTIF is part of the Higher-Order Cognitive Factors Battery (BAFACALO), which was created by C. M. A. Gomes after his doctorate studying the Carroll's model of intelligence (Gomes, 2005; Gomes & Borges, 2007, 2008b). BAFACALO measures cognitive abilities of the Cattell-Horn-Carroll model (Golino & Gomes, 2014a, 2014b) and it has 18 intelligence tests. The tests are available only for research and teaching purposes, on Researchgate platform (Gomes & Nascimento, 2021a, 2021b, 2021c, 2021d, 2021e, 2021f, 2021g, 2021i, 2021j, 2021l, 2021m, 2021n, 2021o, 2021p; Gomes, Nascimento, et al., 2021a, 2021b, 2021c, 2021d). BAFACALO has evidence of internal validity (Gomes, 2010b, 2011b, 2012; Gomes & Borges, 2009a, 2009b, 2009c; Gomes, de Araújo, et al., 2014; Gomes & Golino, 2015) and external validity (Alves et al., 2012; Gomes, 2010a; Gomes & Golino, 2012a, 2012b; Gomes, Golino, et al., 2014). BAFACALO is a reference for the construction of many other intelligence tests, such as Inductive Reasoning Development Test [Logical Reasoning Development Test (TDRI)] (Golino & Gomes, 2015, 2019; Golino, Gomes, et al., 2014) and TDRI-SR (Gomes, Araujo, et al., 2021).
Approaches to Learning Scale (EABAP)	The EABAP comes from a line of research on students' beliefs about teaching and learning (Gomes & Borges, 2008a). It is a self-report test and has 9 items for the deep approach measure and 8 items for the surface approach measure. The EABAP has several pieces of evidence about its internal and external validity (Gomes, 2010c, 2011a, 2013; Gomes, Araujo, et al., 2020; Gomes & Golino, 2012b; Gomes et al., 2011; Gomes, Farias, et al., 2022), as well as being a reference for the construction of other tests of learning approaches (Araujo et al., 2023; Carvalho & Gomes, 2023; Gomes, 2021, 2022; Gomes, Araujo, et al., 2022; Gomes, Jelihovschi, et al., 2022; Gomes & Linhares, 2018; Gomes, Linhares, et al., 2021; Gomes & Nascimento, 2021h, 2021k; Gomes, Quadros, et al., 2020; Rodrigues & Gomes, 2022; Santos et al., 2023).
Booklets for Testing Academic Knowledge and Metacognition (CTCAM-Monitoring)	The CTCAM is composed of three booklets aiming to measure the following constructs: academic knowledge, monitoring, and judgment (Costa, 2018). Each booklet has 40 items, 10 of them to measure academic knowledge, 10 to measure monitoring, and 20 to measure judgment. The booklets have some items in common. In our empirical analysis example, we used only the common items that were answered by all participants pertaining to measure the monitoring, i.e., items 4, 5, 8, and 10. Monitoring is the metacognitive ability of people to detect errors at the moment they are performing a task/activity. Validity evidence and more details about the Academic Knowledge and Metacognition Testing Booklets are presented in Costa (2018).

Note. The data that support the examples of this study are available from the corresponding author upon request.

6. Conclusion

The methodology presented in the article contributes to the methodological systematization of Grice's (2001) proposal and also creates a methodology that allows factor scores to be evaluated in the context of various measurement models, the greatest contribution of our work.

We refine the correlational accuracy criterion of Grice (2001) by creating two categories of analysis: internal and external correlational accuracy. These categories highlight two distinct contexts in which factor scores should be evaluated. Internal correlational accuracy represents the context of the example presented by Grice (2001), i.e., an evaluation of the factor scores extracted from a single measurement model. On the other hand, external correlational accuracy indicates the context in which the factor scores evaluated come from different measurement models. Each context demands its own evaluation, since factor scores can present adequate internal correlational accuracy and inadequate external correlational accuracy, and vice versa.

We present a step-by-step methodological procedure for the execution of the evaluation of internal correlational accuracy and external correlational accuracy. It presents objective processes that allow the researcher to use a well-defined and executable procedure. To execute it, the researcher only needs to have basic knowledge of confirmatory factor analysis or structural equations modeling.

We hope that our article will encourage the scientific community to routinely evaluate the correlational accuracy of factor scores, especially if these scores are used for analyses. As we argue, a relevant bias in correlational accuracy substantially compromises the quality of the measures and, consequently, the quality of the analyses that use them.

7. References

Alves, F. A., Flores, R. P., Gomes, C. M. A., Golino, H. F. (2012). Preditores do rendimento escolar: Inteligência geral e crenças sobre ensino-aprendizagem [Predictors of school performance: General intelligence and beliefs about teaching and learning]. *Revista E-PSI, 1*, 97-117. https://revistaepsi.com/artigo/2012-ano2-volume1-artigo5/

Araujo, J., de Souza Daniel, M., & Gomes, C. (2023). The correction guide of the Approach-In-Process Test (version 2) and its use in the content “Epistemological issues in knowledge acquisition in the theory of Jean Piaget”. *European Journal of Education Studies*, *10*(4), 29-58. http://dx.doi.org/10.46827/ejes.v10i4.4741

Beauducel, A., Hilger, N., & Kuhl, T. (2023). The trade-off between factor score determinacy and the preservation of inter-factor correlations. *Educational and Psychological Measurement*, 1-25. https://doi.org/10.1177/00131644231171137

Costa, B. C. G. (2018). *Caminhos para predição do desempenho acadêmico: um modelo de variáveis cognitivas e socioemocionais* [Pathways to predict academic performance: a model of cognitive and socio-emotional variables]. [Doctoral dissertation, Universidade de Brasília]. https://repositorio.unb.br/handle/10482/32994

Croon M. (2002). Using predicted latent scores in general latent structure models. In Marcoulides G., Moustaki I. (Eds.), *Latent variable and latent structure modeling* (pp. 195-223). Mahwah, NJ: Lawrence Erlbaum

Carvalho, J., & Gomes, C. (2023). Applying the correction guide of the Approach-In-Process Test (version 2) in the electrical current content. *European Journal of Education Studies, 10*(4), 109-142. http://dx.doi.org/10.46827/ejes.v10i4.4753

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis Testing Using Factor Score Regression. *Educational and Psychological Measurement*, *76*(5), 741–770. https://doi.org/10.1177/0013164415607618

Devlieger, I., Talloen, W., & Rosseel, Y. (2019). New developments in factor score regression: Fit indices and a model comparison test. *Educational and Psychological Measurement*, *79*(6), 1-12. https://doi.org/10.1177/0013164419844552

Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement*, *78*(5), 762-780. https://doi.org/10.1177/0013164417719308

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. https://revistaepsi.com/artigo/2014-ano4-volume1-artigo4/

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), e6. https://doi.org/10.5334/jopd.af

Golino, H. F., & Gomes, C. M. A. (2015). Investigando estágios de desenvolvimento do raciocínio indutivo usando a análise fatorial confirmatória, o modelo logístico simples de Rasch e o modelo de teste logístico linear (Rasch estendido) [Investigating developmental stages of inductive reasoning using confirmatory factor analysis, the Rasch simple logistic model and the linear logistic test model (extended Rasch)]. In Hudson F. Golino et al., *Psicometria contemporânea: compreendendo os Modelos Rasch* (pp. 283-338). São Paulo: Casa do Psicólogo. ISBN: 97885845989

Golino, H. F. & Gomes, C. M. A. (2019) *TDRI: Teste de Desenvolvimento do Raciocínio Indutivo* [Inductive Reasoning Developmental Test (IRDT)]. São Paulo: Hogrefe.

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. (2005). *Uma análise dos fatores cognitivos mensurados pelo Exame Nacional do Ensino Médio (ENEM)* [An analysis of cognitive factors measured by the National High School Exam (ENEM)]. [Doctoral dissertation, Programa de Pós-Graduação em Educação da Universidade Federal de Minas Gerais]. http://hdl.handle.net/1843/FAEC-85RJNN

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

Gomes, C. M. A. (2010b). Estrutura fatorial da Bateria de Fatores Cognitivos de Alta-Ordem (BaFaCalo) [Factorial structure of Higher-Order Cognitive Factors Kit]. *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 [Students' profiles and the relationship between learning approach and achievement]. *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 [Deep and surface approach to learning: different perspectives about academic achievement]. *Psicologia: Reflexão e Crítica*, *24*(3), 438-447. https://www.scielo.br/j/prc/a/J6MjLPnqWpQLdt9DRqjPqZB/?lang=pt

Gomes, C. M. A. (2011b). Validade do conjunto de testes da habilidade de memória de curto-prazo (CTMC) [Short term memory ability tests kit validity (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) [Construct validity of the set of crystallized intelligence tests from higher-order cognitive factors kit]. *Gerais: 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 [The construction of a measure of learning approaches]. *Psico (PUCRS. Online), 44*(2), 193-203. http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/11371

Gomes, C. M. A. (2021). Avaliação educacional focada no processo: apresentando o teste SLAT-Thinking 2 [Process-focused educational assessment: Introducing the SLAT-Thinking 2 test]. *Conference. XVI Congresso Internacional Galego-Português de Psicopedagogia*. https://doi.org/10.13140/RG.2.2.24903.42408

Gomes, C. M. A. (2022). Teste Abordagem-em-Processo (Versão 2) [Approach-in-Process Test (Version 2)]. *Preprint*. https://doi.org/10.13140/RG.2.2.29156.24962

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

Gomes, C. M. A., Araujo, J., & Jelihovschi, E. G. (2022). Presentation of the Correction Guide for the Approach-in-Process Test Version 2 and Its Application in the Content of “We Don’t Have Direct Access to Reality. *European Journal of Education and Pedagogy*, *3*(6), 112-123. https://doi.org/10.24018/ejedu.2022.3.6.497

Gomes, C. M. A., Araujo, J., Lima, I. P. C., Chaves, V. N. B., & Golino, H. F. (2021). Inductive Reasoning Developmental Test – Second Revision (TDRI-SR): content validity. In Ezequiel Martins Ferreira (org.), *A pesquisa em psicologia: contribuições para o debate metodológico* (pp. 36-49). Ponta Grossa: Atena. https://doi.org/10.22533/at.ed.1692115124

Gomes, C. M. A., & Borges, O. N. (2007). Validação do modelo de inteligência de Carroll em uma amostra brasileira [Validation of Carroll intelligence model in one brazilian sample]. *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) [Psychometrical properties of teaching and learning student’s beliefs scale]. *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 [Psychometric proprieties of a set of 45 cognitive tests]. *Fractal: Revista de Psicologia, 20*(1), 195-207. https://doi.org/10.1590/S1984-02922008000100019

Gomes, C. M. A. & Borges, O. N. (2009a). O ENEM é uma avaliação educacional construtivista? Um estudo de validade de construto [Is ENEM a constructivist educational assessment? A study of construct validity]. *Estudos em Avaliação Educacional, 20*(42), 73-87. https://doi.org/10.18222/eae204220092060

Gomes, C. M. A., & Borges, O. N. (2009b). Propriedades psicométricas do conjunto de testes da habilidade visuo espacial [Psychometric properties of visual-spatial ability tests kit]. *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 [Psychometrical properties of fluid intelligence tests kit]. *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., & Golino, H. F. (2012a). O que a inteligência prediz: diferenças individuais ou diferenças no desenvolvimento acadêmico? [What does the intelligence predict: individual differences or academic development differences?] *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) [Incremental validity of the Learning Approaches Scale]. *Psicologia: Reflexão e Crítica*, *25*(4), 400-410. https://doi.org/10.1590/S0102-79722012000400001

Gomes, C. M. A., & Golino, H. F. (2015). A medida de habilidades cognitivas amplas da Bateria de Fatores Cognitivos de Alta Ordem (BAFACALO): empregando o modelo Rasch bifatorial [The measurement of broad cognitive abilities from the Higher Order Cognitive Factors Battery (BAFACALO): employing the bifactor Rasch model]. In Hudson F. Golino et al., *Psicometria contemporânea: compreendendo os Modelos Rasch* (pp. 361-385). São Paulo: Casa do Psicólogo. ISBN: 97885845989

Gomes, C. M. A., Golino, H. F., Pinheiro, C. A. R., Miranda, G. R., & Soares, J. M. T. (2011). Validação da Escala de Abordagens de Aprendizagem (EABAP) em uma amostra Brasileira [Validation of the Learning Approach Scale (LAS) in a Brazilian sample]. *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. https://doi.org/10.9734/BJESBS/2014/10757

Gomes, C. M., Jelihovschi, E. G., & Araujo, J. (2022). Presentation of the Approach-In-Process Test (version 2). *European Journal of Education and Pedagogy*, *3*(4), 81-91. https://doi.org/10.24018/ejedu.2022.3.4.402

Gomes, C. M. A, & Linhares. (2018). Investigação da validade de conteúdo do TAP-Pensamento [Investigating the content validity of SLAT-Thinking]. Poster. *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., Linhares, I. S., Jelihovschi, E. G., & Rodrigues, M. N. S. (2021). Introducing rationality and content validity of SLAT-Thinking. *International Journal of Development Research, 11*(1), 43264-43272. https://doi.org/10.37118/ijdr.20586.01.2021

Gomes, C. M. A., & Nascimento, D. F. (2021a). A medida da habilidade de fluência do modelo CHC: apresentando o Teste de Fluência Ideativa 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [The measurement of the fluency ability of the CHC model: presenting the Ideation Fluency Test 2 of the Higher-Order Cognitive Factors Battery (BAFACALO)]. *Preprint*. https://doi.org/ 10.13140/RG.2.2.35726.28481/2

Gomes, C. M. A., & Nascimento, D. F. (2021b). Acesso aberto ao Teste de Fluência Figural da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) como medida da habilidade ampla de fluência do modelo CHC de inteligência [Open access to the Higher-Order Cognitive Factors Battery Figural Fluency Test (BAFACALO) as a measure of the broad fluency ability of the CHC model of intelligence]. *Preprint*. https://doi.org/10.13140/RG.2.2.15593.62564/1

Gomes, C. M. A., & Nascimento, D. F. (2021c). Acesso aberto e gratuito ao Conjunto de Testes de Inteligência Fluida: Teste de Raciocínio Geral da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Free and open access to the Fluid Intelligence Test Kit: Higher-Order Cognitive Factors Battery General Reasoning Test (BAFACALO).]. *Preprint*. https://doi.org/10.13140/RG.2.2.30509.61921/1

Gomes, C. M. A., & Nascimento, D. F. (2021d). Acesso aberto e gratuito ao Teste de Fluência Ideativa 1 da BAFACALO [Free and open access to the BAFACALO Ideation Fluency Test 1]. *Preprint*. https://doi.org/10.13140/RG.2.2.24821.09442/3

Gomes, C. M. A., & Nascimento, D. F. (2021e). Apresentando o Teste de Flexibilidade de Fechamento da BAFACALO [Introducing the BAFACALO Closing Flexibility Test]. *Preprint*. https://doi.org/10.13140/RG.2.2.31920.28164

Gomes, C. M. A., & Nascimento, D. F. (2021f). Disponibilizando de forma gratuita e aberta o Teste de Memória Associativa 1 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Making the Associative Memory Test 1 of the High-Order Cognitive Factors Battery (BAFACALO) available free of charge]. *Preprint*. https://doi.org/10.13140/RG.2.2.29964.03201/1

Gomes, C. M. A., & Nascimento, D. F. (2021g). Disponibilizando de forma gratuita e aberta o Teste de Velocidade Numérica da BAFACALO [Providing the BAFACALO Numerical Speed Test for free and openly]. *Preprint*. https://doi.org/10.13140/RG.2.2.24114.94407/1

Gomes, C. M. A. & Nascimento, D. F. (2021h). Evidências de validade do Teste de Abordagens de Aprendizagem: Identificação do pensamento contido em textos 2 [Validity evidence of the Learning Approaches Test: Identification of thought contained in texts 2]. *Anais do XVI Congresso Internacional Galego-Português de Psicopedagogia*, 1 a 3 de Setembro de 2021, UMinho, Braga, Portugal (pp. 2426-2438).

Gomes, C. M. A., & Nascimento, D. F. (2021i). Medidas de inteligência cristalizada: disponibilizando o Teste de Compreensão Verbal 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Measures of crystallized intelligence: making available the Verbal Comprehension Test 2 of the High-Order Cognitive Factors Battery (BAFACALO)]. *Preprint*. https://doi.org/10.13140/RG.2.2.36085.09447/1

Gomes, C. M. A., & Nascimento, D. F. (2021j). Medindo a habilidade de rapidez cognitiva do modelo CHC: apresentando o Teste de Velocidade Perceptiva 1 da BAFACALO [Measuring the cognitive quickness ability of the CHC model: Introducing the BAFACALO Perceptual Velocity Test 1]. *Preprint*. https://doi.org/10.13140/RG.2.2.28564.83848/1

Gomes, C. M. A., & Nascimento, D. F. (2021k). Presenting SLAT-Thinking Second Version and its content validity. *International Journal of Development Research, 11*(3), 45590-45596. https://doi.org/10.37118/ijdr.21368.03.2021

Gomes, C. M. A., & Nascimento, D. F. (2021l). Projeto de acesso aberto e gratuito à Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO): o Teste de Compreensão Verbal 1 do Conjunto de Testes de Inteligência Cristalizada [Free and open access project to the High-Order Cognitive Factors Battery (BAFACALO): the Verbal Comprehension Test 1 of the Crystallized Intelligence Test Suite]. *Preprint*. https://doi.org/10.13140/RG.2.2.22663.32165/1

Gomes, C. M. A., & Nascimento, D. F. (2021m). Projeto de acesso aberto e gratuito aos testes do LAICO: Teste de Raciocínio Lógico da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Open and free access project to LAICO tests: High-Order Cognitive Factors Battery Logical Reasoning Test (BAFACALO)]. *Preprint*. https://doi.org/10.13140/RG.2.2.25476.45445/1

Gomes, C. M. A., & Nascimento, D. F. (2021n). Projeto de acesso aos testes de inteligência da BAFACALO: Teste de Compreensão Verbal 3 [BAFACALO Intelligence Test Access Project: Verbal Comprehension Test 3]. *Preprint*. https://doi.org/10.13140/RG.2.2.10499.84001/2

Gomes, C. M. A., & Nascimento, D. F. (2021o). Projeto de acesso da BAFACALO: Teste de Memória Associativa 2 [BAFACALO Access Project: Associative Memory Test 2]. *Preprint*. https://doi.org/10.13140/RG.2.2.23253.14565/1

Gomes, C. M. A., & Nascimento, D. F. (2021p). Teste de Memória Visual da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Visual Memory Test of the High-Order Cognitive Factors Battery (BAFACALO)]. *Preprint*. https://doi.org/10.13140/RG.2.2.33319.47529

Gomes, C. M. A., & Nascimento, D. F., & Araujo, J. (2021a). Acesso aberto ao Teste de Dobraduras (VZ) da BAFACALO [Open access to the BAFACALO Folding Test (VZ)]. *Preprint*. https://doi.org/10.13140/RG.2.2.21853.95201/2

Gomes, C. M. A., Nascimento, D. F., & Araujo, J. (2021b). Medindo a inteligência fluida: o Teste de Indução da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Measuring fluid intelligence: the High-Order Cognitive Factors Battery Induction Test (BAFACALO)]. *Preprint*. https://doi.org/10.13140/RG.2.2.17087.84641/3

Gomes, C. M. A., & Nascimento, D. F., & Araujo, J. (2021c). Projeto de testes gratuitos e abertos do LAICO: Teste de Velocidade Perceptiva 3 da BAFACALO [LAICO's free and open testing project: BAFACALO Perceptual Velocity Test 3]. *Preprint*. https://doi.org/10.13140/RG.2.2.36278.42563/2

Gomes, C. M. A., & Nascimento, D. F., & Araujo, J. (2021d). Teste de Velocidade Perceptiva 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO): disponibilização aberta e gratuita aos testes de medida de rapidez cognitiva do LAICO [Perceptual Speed Test 2 of the High-Order Cognitive Factors Battery (BAFACALO): open and free availability to LAICO cognitive quickness tests]. *Preprint*. https://doi.org/10.13140/RG.2.2.29567.53928/2

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*, *25*(1), 33-43. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-294X2020000100004&lng=pt&nrm=iso&tlng=en

Grice, J. W. (2001) Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430-450. https://doi.org/10.1037/1082-989X.6.4.430

Jobst, L. J., Bader, M., & Moshagen, M. (2023). A tutorial on assessing statistical power and determining sample size for structural equation models. *Psychological Methods*, *28*(1), 207–221. https://doi.org/10.1037/met0000423

McDonald, R. P. (1981). Constrained least squares estimators of oblique common factors. *Psychometrika*, *46*(3), 337-341. https://doi.org/10.1007/BF02293740

Rodrigues, M., & Gomes, C. (2022). The correction guide of the Approach-In-Process Test version 2 is a tool for reflection of the pedagogical practice. *European Journal of Alternative Education Studies*, *8*(1), 1-29. http://dx.doi.org/10.46827/ejae.v8i1.4598

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Santos, A., Araujo, J., & Gomes, C. (2023). Applying the correction guide for the Approach-In-Process Test (version 2) to the content "adolescence as social construction". *European Journal of Education Studies*, *10*(5), 57-84. http://dx.doi.org/10.46827/ejes.v10i5.4774

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*(4), 563-575. https://doi.org/10.1007/BF02296196

Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's some interesting parallels. *Psychometrika*, *44* (1), 157-167. http://doi.org/10.1007/BF02293967

Ten Berge, J. M., Krijnen, W. P., Wansbeek, T., & Shapiro, A. (1999). Some new results on correlation-preserving factor scores prediction methods. *Linear algebra and its applications*, *289*(1-3), 311-318. https://doi.org/10.1016/S0024-3795(97)10007-6

Thoemmes, F., Rosseel, Y., & Textor, J. (2018). Local fit evaluation of structural equation models using graphical criteria. *Psychological Methods*, *23*(1), 27–41. https://doi.org/10.1037/met0000147