

## Rápido ou Devagar? Evidências de Validade do Teste de Reflexão Cognitiva (CRT) para o Brasil

Thiago Augusto Costa de Olival<sup>5</sup>, Elaine Rabelo Neiva<sup>6</sup>

*Universidade de Brasília - UNB*

### Resumo

Uma das maneiras utilizadas para medir racionalidade é com o instrumento chamado *Cognitive Reflection Test* (CRT) que possui 3 itens e objetiva a deliberação sobre problemas de cunho lógico matemático envolvendo a interpretação do problema e a realização de uma expressão algébrica simples. A medida sugere que quanto mais acertos, maior o traço latente (mais racional ou maior tendência a ser racional) e quanto mais erros, menor a tendência a racionalidade e em contrapartida, maior a tendência a intuição. Com o objetivo de avaliar os indícios de validade do instrumento para o contexto brasileiro, apresentam-se as análises e resultados decorrentes do processo de adaptação do instrumento. Em uma amostra total de 662 indivíduos foram conduzidas a análise fatorial exploratória, além da realização da TRI (Teoria de Resposta ao Item).

**Palavras-chave:** Racionalidade, Teoria do sistema dual, Validade e confiabilidade

<sup>5</sup> Doutorando no Programa de Psicologia Social, Trabalho e Organizações da Universidade de Brasília, UNB  
Email: thiagoolival@hotmail.com, Telefone: +55 11 99699-1406

<sup>6</sup> Docente no Programa de Psicologia Social, Trabalho e Organizações da Universidade de Brasília, UNB  
Email: elaine.neiva@gmail.com

## Resumen

Una de las formas utilizadas para medir la racionalidad es mediante el instrumento llamado Cognitive Reflection Test (CRT), que consta de 3 elementos y tiene como objetivo la deliberación sobre problemas de lógica matemática que involucran la interpretación del problema y la realización de una expresión algebraica simple. La medida sugiere que cuanto más aciertos, mayor es la tendencia latente a ser más racional y cuanto más errores, menor es la tendencia a la racionalidad y, por el contrario, mayor es la tendencia a la intuición. Con el objetivo de evaluar los indicios de validez del instrumento para el contexto brasileño, se presentan los análisis y resultados derivados del proceso de adaptación del instrumento. En una muestra total de 662 individuos se llevaron a cabo el análisis factorial exploratorio, así como la realización de la Teoría de Respuesta al Item (TRI).

**Palabras clave:** Racionalidad, teoría del sistema dual, validez y confiabilidad.

## Introdução

Com apenas 3 itens, a proposta do Teste de Reflexão Cognitiva (CRT) é de mensurar os fenômenos racionalidade e intuitividade, ambos inseridos na teoria do sistema dual de processamento (Evans & Stanovich, 2013). O construto em questão é baseado na hipótese de que quanto maior a pontuação no teste, mais racional o indivíduo seria. Em contrapartida, quanto menor a pontuação, maior a probabilidade de classificação do julgamento realizado como intuitivo. Apesar de pre-

sente na literatura não acadêmica, sendo encontrada em livros ou na internet, Pennycook e Rand (2019) apresentam evidências de que a divulgação do instrumento não comprometeria sua validade preditiva (Bialek & Pennycook, 2018; Meyer et al., 2018; Staganaro et al., 2018; Srol, 2018).

Neste artigo apresentam-se a adaptação do instrumento para o Brasil, algumas observações de indícios de validade transcultural, de conteúdo e de construto do teste, além de avaliar a dificuldade e o poder de discriminação dos itens.

Os itens presentes no instrumento são de avaliação lógica, e objetivam a interpretação da sentença e a solução de um problema de base algébrica. A suposição que envolve o erro no teste é que, ao se deparar com uma das questões, o indivíduo formule uma resposta que poderia ser equivocada e baseada em uma interpretação insuficiente do enunciado.

O teste tem sido utilizado repetidamente em pesquisas sobre decisão, mostrando seu poder preditivo em temas relacionados a tomada de decisão financeira, avaliação probabilística e boas escolhas de investimento (Neyse, 2016; Alós-Ferrer & Hügelschäfer, 2016; Pennycook, 2018). O grande objetivo do instrumento é que ele poderia distinguir as pessoas que raciocinaram melhor e deram a resposta correta, mas ainda permanecem dúvidas de que as respostas erradas sejam erros intuitivos ou matemáticos. A suposta falta ou excesso de “intuitividade” neste domínio, conforme a hipótese de Pennycook et al. (2016), pode ocorrer em virtude da experiência com problemas matemáticos que alguns indivíduos possuem e não porque são mais ou menos intuitivos.

## Método

### Adaptação do teste ao contexto brasileiro

Com o intuito de evitar que os instrumentos sejam traduzidos de forma literal e que tal processo ocasione um colapso dos indícios de validade do construto, Sireci (2005) sugere que a adaptação de instrumentos entre as culturas seja realizada de maneira a considerar que os conceitos atribuídos sejam preservados.

Alguns dos pressupostos encontrados na literatura sobre a tradução e adaptação do instrumento e que foram atendidos plenamente por este estudo, envolvem: a utilização de dois tradutores bilíngues e experientes (Cassepp-Borges et al., 2010); comparação com a versão traduzida do livro Rápido e Devagar encontrada no Brasil; e a consideração de que os tradutores não possuam familiaridade com o construto envolvido (Beaton et al., 2000). A versão de consenso que atendia aos itens observados por Borsa et al. (2012) apresentou algumas alterações semânticas e de tradução para melhorar a adaptação ao contexto brasileiro.

O grupo de juízes foi composto 8 juízes com formação em Psicologia, Economia ou Administração

de Empresas, mestrandos ou doutorandos. A taxa de concordância observada pelo comitê foi superior aos 90% indicados por Alexandre e Coluci (2011), considerando a proporção de itens que receberam notas 4 ou 5 em uma escala de 1 a 5, em que o 5 representa o maior grau de concordância quanto a clareza e relevância dos itens e 1 referindo-se ao menor grau.

Muito mais do que a tradução literal já exposta anteriormente como insuficiente para garantir a qualidade integral da adaptação, Beaton et al. (2000), Sireci et al. (2005) e Borsa et al. (2012) recomendam que seja considerada a possibilidade de verificar se o contexto do instrumento original permanece, em um processo de tradução reversa. Desta forma, o instrumento foi submetido a dois universitários nativos em língua inglesa residentes nos Estados Unidos e que possuem o português como segunda língua. A versão final do instrumento permaneceu conforme abaixo:

1. Um taco e uma bola custam RS 1,10 no total. O taco custa um real a mais que a bola. Quanto custa a bola?
2. Se 5 máquinas demoram 5 minutos para fazer 5 ferramentas, quanto tempo leva para 100 máquinas fabricarem 100 ferramentas?
3. Em um lago, há um ramo de lírios. Todos os dias, o ramo dobra de tamanho. Se leva 48 dias para o ramo cobrir o lago todo, quanto tempo levará para que o ramo cubra metade do lago?

### Participantes

O instrumento foi aplicado entre o final do ano de 2019 e início do ano de 2020, via *google forms* em comunidades de uma rede social. O processo de aplicação e coleta de dados totalizou 662 observações. Tendo em vista que a pesquisa possuía *link* estático, não é possível saber quantas pessoas de fato acessaram sem concluir o questionário. Os dados da amostra estão apresentados na Tabela 1.

Tabela 1

Caracterização dos participantes da pesquisa

Características	Distribuição (n = 662)
<b>Sexo</b>	
Masculino	293 (44)
Feminino	369 (56)
<b>Escolaridade</b>	
Nível Médio	75 (11)
Superior-Graduação	207 (31)
Superior-Pós-Graduação -Especialização	332 (50)
Superior – Mestrado	39 (6)
Superior - Doutorado	9 (1)
<b>Idade</b>	
18-25	40 (6)
26-30	57 (9)
31-35	121 (18)
36-40	175 (26)
41-45	107 (16)
46-50	68 (10)
51-55	45 (7)
56-60	27 (4)
61-65	14 (2)
66-70	6 (1)
71-75	2 (0)
<b>Região</b>	
Centro Oeste	162 (24)
Nordeste	34 (5)
Norte	13 (2)
Sudeste	399 (60)
Sul	54 (8)

### Análise dos Dados

Os dados avaliados da amostra não apresentaram normalidade, além da inexistência de casos extremos univariados ou dados omissos. Também não

foram observados casos extremos multivariados ao se analisar a distância de *Mahalanobis*. Embora exista a preocupação no que tange aos pressu-

postos, as recomendações de Hair et al, (2009) são de que o tamanho da amostra ( $n > 200$ ) e a inexistência de heterocedasticidade possam ser indicativos de que a não normalidade univariada dos dados não acarrete maiores problemas para o prosseguimento das análises.

## Resultados

Os indícios de validade interna foram analisados por meio dos resultados da análise fatorial, bem como os indícios de validade convergente e discriminante (por meio da confiabilidade composta e da variância média extraída), além dos resultados da TRI.

## Análise Fatorial

Uma análise fatorial exploratória para avaliar a consistência interna de acordo com pressupostos psicométricos (Hair et al., 2009; Pasquali, 2010; Damásio, 2012; Borsa et al. 2012; Tabachnick & Fidel, 2014) foi realizada em uma amostra com 662 indivíduos foi conduzida pelo executável *Factor* versão 10.10.03 com o intuito de avaliar a fatorialidade do instrumento, o que resultou em um KMO limítrofe para que seja considerado como de

razoável a bom ( $KMO=0,72$ ) sugerido por Hutcheson e Sofroniou (1999) além da significância no teste de esfericidade ( $\chi^2=868,1, p>0,00$ ).

O modelo apresentou uma estrutura platicúrtica nos 3 itens ( $q1=-1,76$ ;  $q2=-1,99$ ;  $q3=-1,97$ ) e assimetria levemente inclinada à direita ( $q1=0,50$ ;  $q2=-0,08$ ;  $q3=0,16$ ). Os fatores foram extraídos utilizando os métodos *Robust Diagonally Weighted Least Squares* e *Minimum Rank Factor Analysis* via análise paralela (Damásio, 2012; Machado et al., 2014; Gomes et al., 2016) produzindo resultados semelhantes.

A partir da análise preliminar, a proporção da variância explicada pelo instrumento foi de 0,77, com apenas uma dimensão e autovalor acima de 1. As cargas fatoriais observadas ( $q1=0,78$ ;  $q2=0,76$ ;  $q3=0,89$ ) foram elevadas, assim como as comunalidades, que foram superiores a 0,6 e inferiores que 0,9. Sobre a confiabilidade da escala, observou-se um Alfa de Cronbach de 0,71, valor este indicado pela literatura (Hair et al., 2009; Maroco & Garcia-Marques, 2013; Alexandre et al., 2013; Souza et al., 2017) como mínimo aceitável. Os demais índices de ajuste do modelo são apresentados conforme dados transcritos na Tabela 2.

Tabela 2  
Relação dos índices da análise

	Observado	Esperado
$\chi^2$	768,23	-
<i>Gl</i>	3	-
<i>P-valor</i>	0,000	<0,05
$\chi^2/GL$	289,37	<5
<i>CFI</i>	1,00	>0,9
<i>BIC</i>	38,97	-
<i>RMSEA</i>	0,00	<0,08
<i>GFI</i>	1,00	>0,9
<i>CC</i>	0,85	0,70
<i>VME</i>	0,66	0,50

Nota: índices esperados conforme Hair et al (2009)

O ajuste do modelo também demonstrou indicies satisfatórios (*RMSEA* = 0; *CFI* e *TLI* = 1), o que estaria de acordo com os parâmetros sugeridos, porém podem indicar um superajuste devido a influência da quantidade reduzida de itens do instrumento (Hu & Bentler, 1999; Kenny & McCoach, 2003). A correlação média ao quadrado entre os itens foi de 0,43, sendo menor do que a variância média extraída (*VME* = 0,66), indicando uma evidência de validade discriminante.

### TRI (Teoria de Resposta ao Item)

O modelo de 2PL foi assumido para fins de conhecimento da dificuldade e discriminação dos itens, sendo acerto ao acaso (c) desprezado. Para realizar a análise, as respostas foram transformadas em dados binários, ocasionando em 8 padrões de resposta possíveis, considerando uma matriz composta da quantidade total de itens corretos e

incorretos e quais itens compõem o padrão (apenas o item 1 correto e demais incorretos, item 1 e 3 corretos e item 2 incorreto etc.).

Quando avaliada a correlação entre os itens considerando a técnica da TRI, observa-se que individualmente os itens têm grande índice de correlação bisserial variando de 0,78 a 0,82, sendo desejável

acima de 0,30 (Pasquali, 2018). Ademais, o instrumento apresentou um índice de informação de 7,51 (99,73% considerando -3, 3) para um  $AIC = 2382,54$ ,  $BIC = 2409,51$  e  $-2 \text{ Log Likelihood} = -1185,27$ , considerando o modelo com três itens como o mais adequado e na tabela 3 são apresentados os demais índices do instrumento.

Tabela 3

Índices dos itens do modelo

Variável	$\alpha$	$a$	$b$	Acerto	Erro	$M$	$SD$	$\chi^2$
Q1	0,64	2,110	0,393	38%	62%	0,38	0,49	358,89
Q2	0,65	1,965	-0,065	52%	48%	0,52	0,50	385,38
Q3	0,56	3,456	0,111	46%	54%	0,46	0,50	176,01

Nota:  $a$  = dificuldade;  $b$  = discriminação; Medidas do  $\alpha$  da escala considerando a exclusão do item e  $\chi^2$  considerando  $p > 0,001$ .

Comparando com o modelo de 1PL, observa-se que o  $AIC$  (2383,89 para 1PL) e  $-2 \text{ Log Likelihood}$  (-1189,43 para 1PL) são maiores e o  $BIC$  (2404,84 para 1PL) é menor, mas os três índices encontram-se muito próximos ao resultado observado no modelo de 2PL. Por meio da análise do qui quadrado com base no  $-2 \text{ Log Likelihood}$ , observou-se que as diferenças são significantes, e

que o modelo 2PL apresenta melhoria na análise dos dados.

### Comparação Entre Resultados Brasileiros e Estudos de Outros Países

Os resultados da aplicação do instrumento no Brasil apresentaram consideráveis semelhanças com resultados encontrados em estudos internacionais no que tange a estrutura fatorial, sua



carga e outros indicadores. As médias e o desvio padrão por item convergem com as encontradas pelo estudo de Campitelli e Gerrans (2014) e Kirkegaard e Nordbjerb (2015). Na tabela 4 são

comparados os resultados de Frederick (2005), Kirkegaard e Nordbjerb (2015) e os observados no presente estudo no que tange as médias e padrão de escore.

Tabela 4

*Comparativo Entre Estudos*

Local da coleta	M	n	% de acertos			
			0	1	2	3
MIT	2,18	61	7%	16%	30%	48%
Princeton University	1,63	121	18%	27%	28%	26%
Boston Fireworks Display	1,53	195	24%	24%	26%	26%
Carnegie Mellon Univ	1,51	746	25%	25%	25%	25%
Harvard University	1,43	51	20%	37%	24%	20%
Univ. of Michigan: Ann Arbor	1,18	1267	31%	33%	23%	14%
Estudos na internet	1,10	525	39%	25%	22%	13%
Bowling Green University	0,87	52	50%	25%	13%	12%
Univ. of Michigan: Dearborn	0,83	154	51%	22%	21%	6%
Michigan State Univ.	0,79	118	49%	29%	16%	6%
Universisty of Toledo	0,57	138	64%	21%	10%	5%
Total (Frederick (2005)	1,24	3428	33%	28%	23%	17%
Kirkegaard & Nordjerb (2015)	1,11	72	-	-	-	-
Amostra deste estudo	1,36	662	34%	21%	21%	24%

*Nota.* Considerando o objetivo do presente estudo, é importante ressaltar que o instrumento apresentou resultados similares aqueles encontrados em outros países (ver Frederick ,2005; Campitelli & Gerrans,2014; Kirkgaard & Nordbjerd, 2015).

A comparação com um modelo de 2PL via TRI aplicado por Primi et al., (2016) também apresentou resultados similares ao deste estudo, em que se observou que os três itens possuem discrimina-

ção alta ( $q1_{\sigma}=1,7$ ;  $q2_{\sigma}=1,48$ ;  $q3_{\sigma}=3,05$ ) e dificuldade de fácil a média ( $q1_b=0,26$ ;  $q2_b =0,70$ ;  $q3_b=0,08$ ) para um índice de informação de 6,10.

## Discussão

Ao que se demonstra, existem diversas similaridades dos resultados observados neste estudo quando comparados com os achados de estudos internacionais (ver Campitelli & Gerrans, 2014; Kirkegarrd & Nordbjerb, 2015). O processo de adaptação ao contexto brasileiro foi realizado em conformidade com os pressupostos elencados (Beaton et al., 2000; Sireci et al., 2005; Borsa et al., 2012; Cassep-Borges et al., 2010). Desta forma, sugere-se que a adaptação e tradução permitiram avaliar indícios de validade transcultural com muita pertinência. Tais resultados apresentam evidências de que o processo da adaptação transcultural não alterou a essência do instrumento. Considera-se que existem indícios para subsidiar a utilização do instrumento no contexto brasileiro.

Existem indícios de validade de conteúdo observados em consonância com as sugestões de Alexandre e Coluci (2011), visto que o grau de convergência entre as análises de 8 juízes e especialistas foi elevada e o  $IVC = 100\%$ , considerando uma escala de 1 a 5. O cálculo proposto é representado pela somatória de respostas com escore 4 ou 5 e dividido pelo número total de respostas. Embora este cálculo seja costumeiramente apre-

sentado com escalas de 1 a 4, Alexandre e Coluci (2011) não sugerem óbices quanto a utilização da escala com 5 pontos. Os parâmetros avaliados foram clareza, pertinência e relevância dos itens.

A confiabilidade do instrumento ( $\alpha = 0,71$ ;  $\lambda 2 = 0,72$ ) também mantém média semelhante aos estudos internacionais, embora ainda seja motivo de atenção pois é considerada como próxima ao limite considerável como aceitável para os pressupostos gerais de confiabilidade, observada em múltiplos artigos sobre o tema (Maroco & Garcia-Marques, 2013; Souza et al, 2017). A confiabilidade composta está adequada as indicações de Hair et al., (2009) e Valentini e Damasio (2016), pois apresentam valor de 0,85. Ao mesmo tempo, a VME entre os itens e o fator também se encontra superior ao valor do pressuposto, o que sugere evidências de validade convergente. Ainda, observam-se evidências de validade discriminante visto que a VME é maior do que as correlações dos itens ao quadrado. Observa-se que o modelo unifatorial está bem ajustado e com cargas fatoriais robustas para cada um dos itens, o que implicaria em aceitar que apesar de suas limitações em alguns índices, o modelo proposto é adequado, mas com ressalvas.

Uma hipótese para a inadequação do modelo no que tange a relação entre os graus de liberdade e  $X^2$  pode ser devido a quantidade de itens presentes no instrumento (que são apenas 3). A mesma hipótese é discutida por Valentini e Damasio (2016) sobre os possíveis impactos da quantidade de itens nos coeficientes alfa de Cronbach, variância média extraída e confiabilidade composta.

Outro indício de validade sugerido por Pasquali (2018) diz respeito a observação de validade de construto via TRI, visto que com análise é possível verificar o traço latente via função de informação.

As orientações prescritas por Pasquali (2018) quanto a interpretação da TRI, no que diz respeito ao “a” (discriminação), ao “b” (dificuldade) e no que tange a avaliação da curva característica do item, são para considerar a probabilidade de 0,5 para respostas corretas ao acaso. Os itens com “a” acima de 1,70 possuem alta discriminação, isto é, alta capacidade de discriminar sujeitos em relação aos níveis do traço latente avaliado (Andrade et al. 2000) e o que se observou é que todos os itens do instrumento possuem índices satisfatórios sobre este quesito. O item 3 (q3) possui a melhor discriminação, o que demonstra sensibilidade conside-

rável para a mensuração de mesmo nível de *tetha* em múltiplos indivíduos (Sartes & Souza-Formigoni, 2013).

No que tange a dificuldade, ao serem consideradas a expectativa de valores esperados (-2 a +2), conforme Baker (2001), os itens da escala poderiam ser considerados como “medianos”, sendo o item 1 (q1) o mais próximo do parâmetro “difícil” e o item 2 (q2) como o mais próximo de “fácil”, tendo como parâmetro somente a comparação dentro da escala e para esta amostra. Importante ressaltar que assim como no estudo de Primi et al., (2016), a escala apresenta índices satisfatórios para a diferenciação de *tetha* de indivíduos que não estejam entre as extremidades mais baixas ou mais altas do traço latente.

Os índices de dificuldade e discriminação demonstram que os itens possuem boa qualidade para a medição do traço latente de maneira mais central na escala, além de apresentar confiabilidade e estrutura fatorial. Como é uma escala com quantidade reduzida de itens, pode ser facilmente incluída em pesquisas sem que se eleve substancialmente o tempo ou a quantidade de perguntas de um questionário.

É possível concluir que os indícios de validade observados apontam que o construto racionalidade pode ser medido pelo instrumento, mas os resultados sugerem que algumas das ressalvas ou inadequações a pressupostos estatísticos estariam relacionados a quantidade reduzida de itens da escala.

Ao mesmo tempo, permanecem inconclusivas as correlações com a numerácia (Campitelli & Gerrans, 2014) ou outros construtos que tenham como objetivo medidas de racionalidade, pois estas não foram alvo deste estudo. O traço latente observado no teste é sugerido por Frederick (2005) como uma medida da resistência a responder as questões de maneira impulsiva e até o momento não existem evidências que não seja somente uma medida de habilidade aritmética, apesar da relação entre os conceitos.

Como limitação do estudo, destaca-se a possibilidade de existir um viés no que tange ao conhecimento prévio ou a possibilidade de consulta das respostas na internet. Não foi possível saber se os indivíduos realizaram algum tipo de consulta ou se já conheciam as respostas. Para atenuar essa constatação, os resultados e médias observadas foram semelhantes àquelas observadas em

outros estudos. Os possíveis impactos do conhecimento prévio dos itens, caso existam, devem ter se apresentado com o mesmo padrão dos artigos citados neste trabalho.

Em estudos futuros, podem ser feitas propostas de incremento do instrumento utilizando também testes de raciocínio lógico de outras categorias como sequências lógicas ou deduções/ induções, visto que na atual proposta ele contém somente itens de cunho lógico matemático. Uma das hipóteses de agenda futura é que a quantidade de itens poderia influenciar nos indicadores da escala, fato este que tem se mostrado uma tendência em alguns artigos com alguma medida de razão e intuição, visto que os instrumentos têm sido utilizados com a escala original suplementada de itens adicionais de natureza similar. Sugere-se comparar os dados de validade entre versões com três e mais itens.

## Referências

Alexandre, N. M. C., & Colucci, M. Z. O. (2011). Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Ciência & Saúde Coletiva*, 16(7),3061-3068. <https://doi.org/10.1590/S1413-81232011000800006>

- Alós-Ferrer, C., & Hügelschäfer, S. (2015). Faith in Intuition and Cognitive Reflection. *Journal of Behavioral Economics*. 64. 10.1016/j.socec.2015.10.006.
- Andrade, D. F., Valle, R. C., & Tavares, H. R. (2000) *Introdução à teoria da resposta ao ítem: conceitos e aplicações*. Sinapse.
- Baker, F. B. (2001) *The Basics of Item Response Theory*. 2ª. ed. United States of America: ERIC Clearinghouse on Assessment and Evaluation
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191.
- Białek, M., & Pennycook, G. (2018). The Cognitive Reflection Test is robust to multiple exposures. *Behavior Research Methods*. 50. 1953–1959. 10.3758/s13428-017-0963-x.
- Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Adaptação e Validação de Instrumentos Psicológicos entre Culturas: Algumas Considerações. *Paidéia*, 22(53), 423–432. <https://doi.org/10.1590/S0103-863X2012000300014>
- Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Mem Cognit*. 2014 Apr;42(3):434-47. doi: 10.3758/s13421-013-0367-9. PMID: 24132723.
- Cassepp-Borges, V., Balbinotti, M. A. A., & Teodoro, M. L. M. (2010). Tradução e validação de conteúdo: Uma proposta para a adaptação de instrumentos. In L. Pasquali, *Instrumentação psicológica: Fundamentos e práticas*. Porto Alegre: Artmed.
- Couto, G., & Primi, R. (2011). Teoria de resposta ao ítem (TRI): Conceitos elementares dos modelos para itens dicotômicos. *Boletim de Psicologia*, 61(134), 1-15.
- Cueva, C., Iturbe-Ormaetxe, I., Mata-Pérez, E., Ponti, G., Sartarelli, M., Yu, H., & Zhukova, V. (2016). Cognitive (ir)reflection: New experimental evidence. *Journal of Behavioral and Experimental Economics*, 64, 81–93. <https://doi.org/10.1016/j.socec.2015.09.002>
- Damásio, B. F. (2012). Uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica*, 11(2), 213-228.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing

- the Debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Frederick, S. (2005). "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives*, 19(4): 25-42. DOI: 10.1257/089533005775196732
- Gomes, C. A. O., Soares, A. B., Mourão, L., & Hernandez, J. A. E. (2016). Inventário de Estilos de Enamorado: construção e validação. *Avaliação Psicológica*, 15(2), 151-159.
- Hair, Jr., J.F., Willian, B., Babin, B. & Anderson, R.E. *Análise multivariada de dados. 6.ed.* Porto Alegre: Bookman, 2009.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55
- Hutcheson, G. D. & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. London: Sage Publications
- Kahneman, D. (2012). *Rápido e devagar, duas formas de pensar*. Rio de Janeiro: Editora Objetiva.
- Kenny, D., A., & McCoach, D., B., (2003) Effect of the Number of Variables on Measures of Fit in Structural Equation Modelin. *Structural Equation Modeling: A Multidisciplinary Journal*, 10:3, 333-51, DOI: 10.1207/S15328007SEM1003\_1
- Kirkegaard, E. O. W., & Nordbjerg, O. (2015). Validating a Danish translation of the International Cognitive Ability Resource sample test and Cognitive Reflection Test in a student sample. *Open Differential Psychology*. 10.26775/ODP.2015.07.31.
- Machado, W. L., Hauck, N., Teixeira, M. A. P., & Bandeira, D. R. (2014). Análise de teoria de resposta ao item de marcadores reduzidos da personalidade. *Psico*, 45(4), 551-558.
- Maroco, J., & Garcia-Marques, T. (2013). Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas? *Laboratório de Psicologia*. 4. 10.14417/lp.763.
- Meyer, A., & Zhou, E., & Frederick, S. (2018). The non-effects of repeated exposure to the cognitive reflection test. *Judgment and Decision Making*. 13. 246-259.
- Neyse, L., Bosworth, S., Ring, P. et al.(2016). Overconfidence, Incentives and Digit Ratio. *Sci Rep* 6, 23294). <https://doi.org/10.1038/srep23294>

- Pasquali, L. (2010). *Instrumentação Psicológica: fundamentos e práticas*. Porto Alegre: Artmed
- Pasquali, L. (2018) *Teoria de Resposta ao Item – Teoria, procedimentos e aplicações*. 3 ed. Curitiba. Appris
- Pennycook, G. (2018). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*. 188. 10.1016/j.cognition.2018.06.011.
- Pennycook, G., & Rand, D. G. (2019). Cognitive Reflection and the 2016 U.S. Presidential Election. *Personality and Social Psychology Bulletin*, 45(2), 224–239. <https://doi.org/10.1177/0146167218783192>
- Pennycook, G., Cheyne, J.A., Koehler, D.J. et al. (2016) Is the cognitive reflection test a measure of both reflection and intuition?. *Behav Res* 48, 341–348 (2016). <https://doi.org/10.3758/s13428-015-0576>
- Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016) The Development and Testing of a New Version of the Cognitive Reflection Test Applying Item Response Theory (IRT). *J. Behav. Dec. Making*, 29: 453– 469. doi: 10.1002/bdm.1883.
- Sartes, L. M. A., & Souza-Formigoni, M. L. O. (2013). Avanços na psicometria: da Teoria Clássica dos Testes à Teoria de Resposta ao Item. *Psicologia: Reflexão e Crítica*, 26(2), 241-250. <https://doi.org/10.1590/S0102-79722013000200004>
- Sireci, S. G. (2005). Using bilinguals to evaluate the comparability of difference language versions of a test. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 117-138). Mahwah, NJ: Lawrence Erlbaum.
- Souza, A. C., Alexandre, N. M. C., & Guirardello, E. B. (2017). Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde*, 26(3), 649-659. <https://dx.doi.org/10.5123/s1679-49742017000300022>
- Srol, J.(2018). These Problems Sound Familiar to Me: Previous Exposure, Cognitive Reflection Test, and the Moderating Role of Analytic Thinking. *Studia psychologica*. 60. 195-208. 10.21909/sp.2018.03.762.
- Stagnaro, M. N., Pennycook, G., & Rand, D. G. (2018). "Performance on the Cognitive Reflection Test is stable across time," *Judgment and Decision Making, Society for Judgment and Decision Making*, vol. 13(3), pages 260-267, May.

Tabachnick, B. G., and Fidell, L. S. (2014). *Using Multivariate Statistics*, 6th ed. Harlow, UK: Pearson.

Valentini, F, & Damásio, B. F. (2016). Variância Média Extraída e Confiabilidade Composta: Indicadores

de Precisão. *Psicologia: Teoria e Pesquisa*, 32(2), e322225. Epub October 27, 2016. <https://doi.org/10.1590/0102-3772e322225>