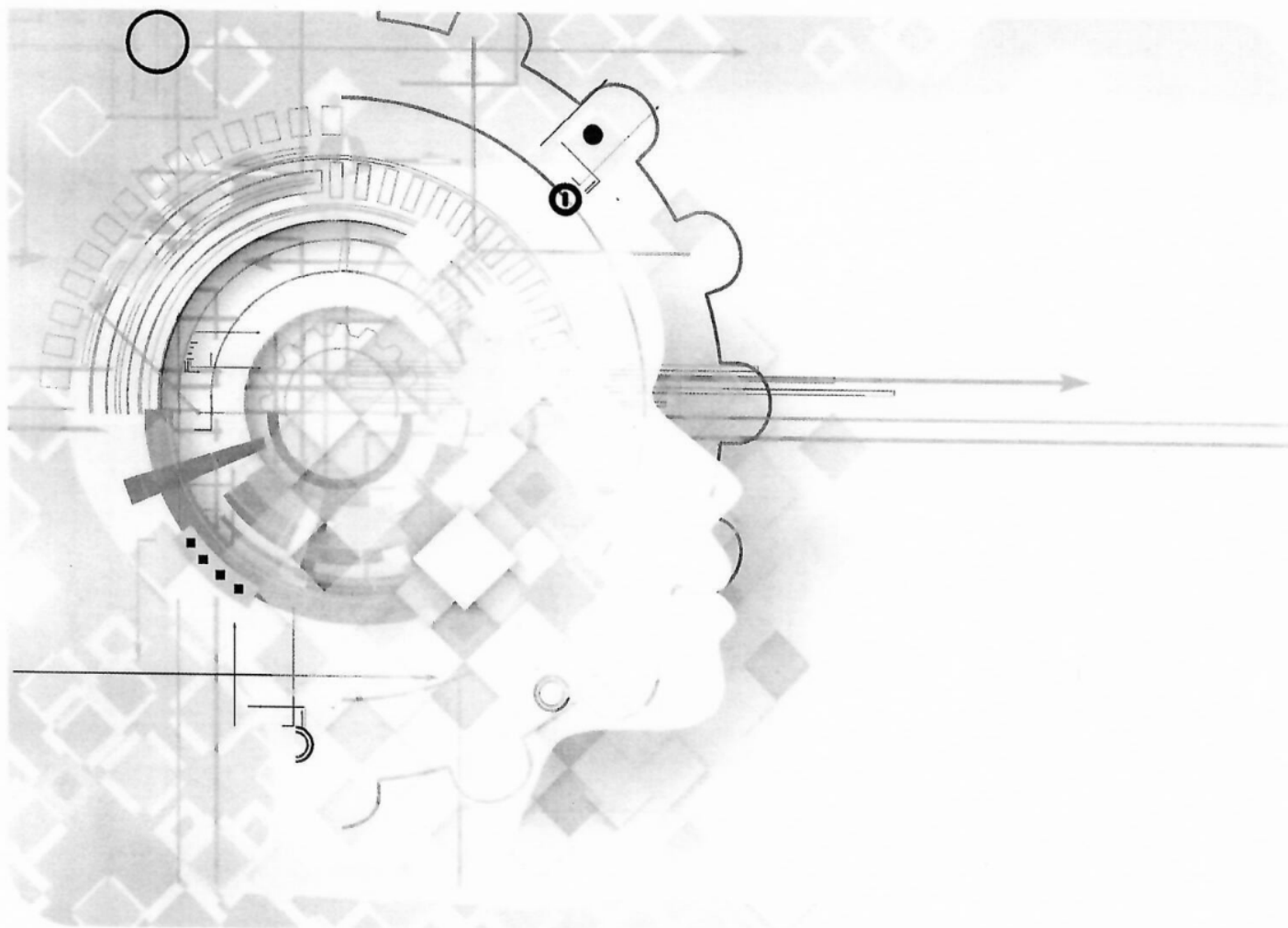


Neuropsicológico



TCS-1/TCS-2

Teste de Cancelamento dos Sinos

Organizadoras

Rochele Paz Fonseca

Caroline de Oliveira Cardoso

Karin Zazo Ortiz

Maria Alice de Mattos Pimenta Parente

Louise Gauthier

Yves Joannette

Livro de instruções Vol. 1

 **Vetor**[®]
editora

5

EVIDÊNCIAS DE FIDEDIGNIDADE DO TESTE DE CANCELAMENTO DOS SINOS – VERSÕES 1 E 2

*Laura Damiani Branco
Cristina Elizabeth Izábal Wong
Caroline de Oliveira Cardoso
Rochele Paz Fonseca*

O presente capítulo tem por objetivo mostrar evidências de fidedignidade das duas versões do TCS-1 e TCS-2. Entre os parâmetros psicométricos que devem guiar a construção e a adaptação de instrumentos, a fidedignidade refere-se à consistência ou precisão do uso do instrumento, na medida em que o teste está livre de erros e/ou que a capacidade de prever o desempenho não varia com o tempo (Pasquali, 2007; Urbina, 2007). Assim, quanto mais parecidos forem os escores dos indivíduos em diferentes momentos ou versões de administração de um instrumento de avaliação, maior será sua fidedignidade (Zanon & Hauck Filho, 2015).

Existem diversas técnicas para a obtenção da fidedignidade, como forma paralela, teste-reteste, duas metades, consistência interna e análise de concordância entre avaliadores a partir do manual de interpretação e pontuação. Dessa forma, buscou-se investigar se há evidências de fidedignidade entre as duas versões do teste por meio de três análises. A primeira conduzida foi a de teste-reteste. Embora haja cada vez mais críticas ao método teste-reteste para testes de atenção (Hedge, Powell & Summer, 2017), em razão da baixa variabilidade dos escores nas tarefas de cancelamento, a técnica, ainda, mais utilizada na obtenção de parâmetros de fidedignidade é a técnica de teste-reteste (Bickerton, Samson, Williamson & Humphreys, 2011; Lee et al., 2004), em que se observa a consistência dos resultados pelo tempo (Pasquali, 2009; Urbina, 2007). Assim, foi realizada a aplicação e reaplicação do teste em uma amostra de participantes saudáveis, procurando-se investigar a estabilidade do desempenho ao longo do tempo. Fez-se, ainda, uma análise de consistência interna entre subescores por coluna e escores totais de ambas as versões do teste (Strauss, Sherman & Spreen, 2006; Urbina, 2014), com o tradicional alfa de Cronbach, acrescentando-se do índice ômega, mais indicado para testes neuropsicológicos (Watkins, 2017).

Por fim, como uma terceira fonte de evidências de fidedignidade, foi realizada uma análise de concordância entre avaliadores, método bastante utilizado para obtenção de parâmetros de fidedignidade em tarefas de cancelamento (Lee et al., 2004; Manly et al., 2009). Neste capítulo, serão expostos, ainda, os resultados de ambos estudos referentes à fidedignidade do TCS-1 e do TCS-2.

5.1 TESTE-RETESTE DOS TCS-1 E TCS-2

5.1.1 Participantes

Participaram do estudo um total de 24 adultos saudáveis, dos quais 58,3% eram do sexo feminino. Ainda, 75% da amostra foi recrutada na cidade de Porto Alegre, RS, enquanto 25% da amostra foi recrutada na cidade de São Paulo, SP. Os participantes tinham entre 23 e 63 anos de idade, e de 5 a 31 anos de escolaridade. Ingressaram na amostra somente aqueles que não apresentaram quaisquer distúrbios sensoriais não corrigidos, sinais sugestivos de quadros neurológicos e/ou psiquiátricos, sinais de depressão e demência, histórico de alcoolismo e relato de uso abusivo atual ou prévio de drogas ilícitas. Os dados de caracterização da amostra podem ser consultados na Tabela 5.1. Para a análise de consistência interna, participaram todos os indivíduos saudáveis da amostra de normatização, caracterizada no Capítulo 11. Participaram desta análise de consistência interna, ainda, os pacientes com acidente vascular cerebral (caracterizados no Capítulo 8) e com traumatismo cranioencefálico (caracterizados no Capítulo 9).

Tabela 5.1. Caracterização sociodemográfica e clínica dos participantes

Variáveis de caracterização da amostra	M	DP
Idade (anos)	38,83	12,96
Escolaridade (anos)	13,58	5,37
Hábitos de leitura e escrita (frequência)	16,58	4,70
Escore socioeconômico	27,79	5,09
MEEM (escore /30)	29,38	1,17
BDI (escore /62)	4,87	4,24

Nota: M = Média, DP = Desvio-padrão; MEEM = Miniexame do Estado Mental, BDI = Inventário de Depressão de Beck.

5.1.2 Procedimentos e instrumentos

De acordo com os aspectos éticos envolvidos em investigações com seres humanos, a participação dos indivíduos foi voluntária, e os instrumentos foram administrados individualmente em um ambiente apropriado. As sessões de avaliação foram realizadas em dois encontros. No primeiro (teste), foram aplicados, além do TCS-1 e TCS-2, os seguintes instrumentos de caracterização da amostra e de critérios de inclusão e exclusão da amostra: Miniexame do Estado Mental (MEEM; Folstein, Folstein, & McHugh, 1975; adaptado por Chaves & Izquierdo, 1992); Inventário de Depressão de Beck (BDI; Beck, Steer, & Brown, 1996, adaptado por Cunha, 2001); questionário sociodemográfico, cultural e clínico (Fonseca et al., 2012). No segundo momento (reteste), houve apenas a aplicação dos TCS-1 e TCS-2, além de uma breve entrevista sobre mudanças na saúde geral e de eventos de vida entre teste e reteste. A sessão de reteste ocorreu entre um e quatro meses após a primeira sessão.

5.1.3 Análise de dados

Os escores de tempo e acurácia obtidos por participante na primeira e na segunda aplicação do TCS-1 e TCS-2 foram investigados, utilizando análises descritivas, seguidos por verificação da normalidade dos dados por meio de teste de Kolmogorov-Smirnov. Após a verificação de que os dados se comportaram de forma não paramétrica, foi realizada a técnica de diferenças entre médias teste Wilcoxon para investigar a significância das diferenças entre as tarefas obtidas no teste e reteste. Análises de consistência interna também foram conduzidas para os subescores correspondentes às quantidades de sinos cancelados por coluna. Isso se deve ao fato de os testes do sinos não serem compostos por itens, mas sim por agrupamentos em colunas que fornecem o escore total. Foram utilizados o alfa de Cronbach e o ômega de McDonald (Revelle & Zinbarg, 2009); para ambos, os

valores considerados adequados seguem as seguintes diretrizes: os coeficientes foram interpretados da seguinte forma: $<0,60$ = inadequados; $0,60$ to $0,69$ = fidedignidade marginal; $0,70$ to $0,79$ = aceitável; $0,80$ to $0,89$ = boa ; e $0,90$ ou mais = excelente (Hair, Black, Babin & Anderson, 2014). As análises foram realizadas com o pacote estatístico SPSS (Statistical Package for the Social Sciences), em sua versão 17.0 e o software R (R Core Team, 2017) a partir das funções implementadas pelo pacote psych (Revelle, 2015).

5.1.4 Resultados e Discussão

A análise descritiva de desempenho da amostra nos momentos teste e reteste de ambas as versões do TCS é apresentada na Tabela 5.2.

Tabela 5.2. Análise comparativa entre escores no teste e reteste dos testes TCS-1 e TCS-2

Variáveis do TCS	Teste M (DP)	Reteste M(DP)	W	p
TCS-1				
Total de omissões T1	1,25 (1,26)	0,71 (1,04)	-1,49	0,13
Total erros (distratores) T1	0,00 (0,00)	0,00 (0,00)		
Tempo total pré-pista - T1 (seg.)	98,48 (37,39)	86,35 (30,94)	-0,80	0,42
Total de omissões T2	0,29 (0,80)	0,08 (0,28)	-1,28	0,19
Total erros (distratores) T2	0,00(0,00)	0,00 (0,00)		
Tempo total pós-pista – T2 (seg.)	57,54 (19,44)	52,50 (20,89)	-0,74	0,45
Tempo Total (seg) T1+T2	156,02 (49,31)	140,94 (52,41)	-0,91	0,36
TCS-2				
Total de omissões T1	1,58 (2,16)	1,25 (3,05)	-1,10	0,26
Total erros (distratores) T1	0,00 (0,00)	0,00 (0,00)	---	
Total erros (distratores sinos sem badalo) T1	0,04 (0,20)	0,04 (0,20)	0	1,00
Tempo total pré-pista – T1 (seg.)	93,17 (28,38)	86,27 (33,55)	-1,18	0,23
Total de omissões T2	0,38 (1,43)	0,75 (1,84)	-2,31	0,02
Total erros (distratores) T2	0,00(0,00)	0,00(0,00)	--	
Total erros (distratores sinos sem badalo) T2	0,00(0,00)	0,00(0,00)	--	
Tempo total pós-pista – T2 (seg.)	51,84 (14,25)	53,46 (25,55)	0	1,00
Tempo Total (seg.) T1 + T2	145,02 (34,69)	139,77 (53,46)	-0,48	0,62

Nota: T1 = Tempo pré-pista; T2 = Tempo pós-pista.

A partir da estatística descritiva apresentada na Tabela 5.2, pode-se observar o quão reduzida é a variabilidade do TCS-1, apenas um pouco maior no TCS-2. Não houve diferenças significativas entre variáveis de acurácia e de tempo entre teste e reteste para todas as variáveis, exceto para total de omissões tempo 2 (pós-pista) do TCS-2. A principal hipótese explicativa é que os participantes devem ter melhorado e aprendido com essa tarefa a partir do monitoramento externo dado pela pista do examinador. De um modo geral, todas as variáveis de tempo 1 (pré-pista) de ambas as versões do teste são as mais importantes clinicamente para o exame dos construtos almejados.

A seguir são apresentados os coeficientes alfa e ômega, que podem ser consultados na Tabela 5.3.

Tabela 5.3. Dados das análises de consistência interna dos TCS-1 e TCS-2

Variáveis	Alfa	Ômega
TCS-1 Total de omissões T1	0,79	0,83
TCS-2 Total de omissões T1	0,75	0,78
TCS-1 Total erros (distratores) T1	0,88	0,99
TCS-2 Total erros (distratores) T1	-	-

5.2 CONCORDÂNCIA ENTRE JUÍZES

5.2.1 Participantes

Participaram do estudo de consistência interna um total de 30 adultos saudáveis, dos quais 70% eram do sexo feminino. Os participantes tinham entre 19 e 75 anos de idade e de 5 a 31 anos de escolaridade. Os critérios de inclusão e exclusão de participantes foram semelhantes aos anteriormente descritos. Os dados de caracterização da amostra podem ser consultados na Tabela 5.4. A análise de juízes foi realizada por duas neuropsicólogas com experiência em avaliação clínica.

Tabela 5.4. Caracterização sociodemográfica e clínica dos participantes

Variáveis de caracterização da amostra	M	DP
Idade (anos)	43,20	16,89
Escolaridade (anos)	15,30	6,05
Hábitos de leitura e escrita (frequência)	18,13	5,48
Escore socioeconômico	27,20	6,23
MEEM (escore /30)	28,73	1,60
BDI (escore /62)	5,67	5,05

Nota: M = Média, DP = Desvio padrão; MEEM = Miniexame do Estado Mental, BDI = Inventário de Depressão de Beck.

5.2.2 Procedimentos e instrumentos

De acordo com os aspectos éticos envolvidos em investigações com seres humanos, a participação dos indivíduos foi voluntária, e os instrumentos foram administrados individualmente em um ambiente apropriado. O procedimento de avaliação, assim como os instrumentos utilizados, foi o mesmo citado na seção anterior deste capítulo.

Após a aplicação e correção dos testes, os protocolos foram entregues para cada uma dos juízes, para a avaliação das estratégias de cancelamento utilizadas por participante. Essa avaliação foi realizada separadamente pelas neuropsicólogas, cuja tarefa consistiu na identificação do uso de estratégias de cancelamento organizadas *versus* desorganizadas, assim como a classificação da estratégia utilizada em uma das seguintes categorias: organizada horizontal começando pela esquerda, organizada horizontal começando pela direita, organizada vertical começando por superior, organizada horizontal começando por inferior, organizada mista (p. ex., horizontal e vertical), organizada outras (p. ex., circular), ou caótica.

5.2.3 Análise de dados

Os julgamentos de cada uma das juízas foram coletados e tabulados separadamente para o TCS-1 e TCS-2. A concordância entre os julgamentos foi investigada por meio da estatística kappa de Cohen e porcentagem de concordância entre juízes. Todas análises foram realizadas com o pacote estatístico SPSS (Statistical Package for Social Science), em sua versão 17.0.

5.2.4 Resultados

Na Tabela 5.5, são apresentados os resultados de concordância entre juízes para o uso de estratégias de cancelamento no TCS-1 e TCS-2. As estratégias de busca utilizadas por participantes foram avaliadas de duas formas distintas. Primeiramente, foram classificadas de forma dicotômica, a fim de distinguir estratégias desorganizadas daquelas com algum nível de organização. Para esse fim, todas as estratégias organizadas – incluindo vertical, horizontal, mista e outras – foram agrupadas

em uma única categoria. O restante foi agrupado em uma segunda categoria, contendo apenas as estratégias desorganizadas, ou caóticas. A concordância dos juízes quanto a essa caracterização dicotômica das variáveis está descrita na primeira linha da Tabela 5.5.

Posteriormente, as estratégias organizadas foram categorizadas de acordo com suas características específicas, de modo que cada participante foi classificado em uma de sete categorias distintas (estratégia desorganizada ou alguma das seis estratégias organizadas anteriormente descritas). A concordância dos juízes quanto a essa caracterização está descrita na segunda linha da Tabela 5.5.

Tabela 5.5. Coeficiente de kappa de concordância entre juízes

Variáveis de caracterização da amostra	K	
	TCS-1	TCS-2
Estratégia organizada vs. desorganizada	0,474	0,651
Tipo de estratégia	0,895	0,946

Nota: T1 = Tempo pré-pista; T2 = Tempo pós-pista;

Verifica-se pela Tabela 5.5 que houve concordância de moderada a forte entre juízes para o tipo de estratégias de cancelamento tanto no TCS-1 quanto no TCS-2. Mais especificamente, a concordância entre juízes para o tipo de estratégia pode ser interpretada como forte para o TCS-1, e quase perfeita para o TCS-2 (McHugh & Barton, 2012). De forma interessante, esses valores foram superiores à concordância entre juízes na classificação de estratégias entre as categorias organizada e desorganizada. Esses achados poderão, em primeira instância, apresentar-se como paradoxais, uma vez que a categorização do “tipo de estratégia” envolve um número muito maior de categorias do que sua classificação entre “organizadas” e “desorganizadas”. Como poderia a concordância ser tão baixa em uma simples categorização dicotômica e tão significativa quando um sistema muito mais complexo de classificação é utilizado?

Esta dúvida pode ser facilmente esclarecida por meio da análise do percentual de concordância entre os juízes. Para complementar a análise de concordância com base no kappa de Cohen, foi realizada avaliação da porcentagem de concordância entre juízes em cada uma das variáveis analisadas. Os resultados desse procedimento estão dispostos na Tabela 5.6.

Tabela 5.6. Análise de concordância entre juízes por percentual

Variáveis de caracterização da amostra	% de concordância	
	TCS-1	TCS-2
Estratégia organizada vs. desorganizada	93,33%	96,67%
Tipo de estratégia	93,33%	96,67%

Como pode ser observado na Tabela 5.6, foi encontrada elevada taxa de concordância entre juízes tanto no TCS-1 quanto no TCS-2, em ambos os sistemas de classificação. A discrepância entre esses achados e aqueles revelados pelo kappa de Cohen pode ser explicada pela diferença entre as duas variáveis analisadas. A primeira contém apenas duas possibilidades de classificação (desorganizada, organizada). A segunda contém sete níveis distintos. Ao computar a concordância entre juízes, a estatística de Cohen considera o número de alternativas disponíveis aos juízes, aferindo um “peso” diferente ao valor final da concordância dependendo do número de alternativas disponíveis aos juízes durante o processo de avaliação. Esse procedimento é realizado para “corrigir” o efeito de diversas fontes de interferência, incluindo as taxas de concordância que possam ocorrer “por acaso”: juízes optando entre apenas duas categorias, por exemplo, poderão concordar em grande parte dos casos se a probabilidade de escolha de cada alternativa for de 50% (caso os desfechos sejam igualmente prováveis). Quando estiverem optando entre três, cinco, dez ou mais categorias, a probabilidade desse tipo de ocorrência é muito menor, de modo que apenas uma pequena parcela

da concordância observada possa ser atribuída ao acaso. Nesses casos, a “correção” necessária no índice de kappa é menor.

Esse procedimento poderá explicar a discrepância entre os valores na Tabela 5.5. A primeira variável, contendo apenas duas alternativas, recebeu um valor de kappa “corrigido”, em que a pequena discordância entre os juízes recebeu um peso muito maior no cálculo estatístico. O oposto ocorre na segunda variável, em que houve um nível similar de discordância, mas valores de kappa muito mais altos. Essa interpretação pode ser corroborada pelos dados na Tabela 5.6, que apresentam elevada concordância entre juízes para ambas variáveis qualitativas do TCS-1 e TCS-2. Ainda, observamos que estas porcentagens são idênticas para as duas variáveis em cada um dos testes. Dessa forma, comprova-se que a diferença entre os valores de kappa nesse caso não se deu por uma real diferença na concordância dos juízes, e sim pela correção estatística realizada.

CONSIDERAÇÕES FINAIS

A partir dos resultados apresentados neste capítulo, observaram-se evidências de fidedignidade das duas versões desta ferramenta na avaliação da atenção e de construtos relacionados por meio de três fontes distintas de análises. Houve estabilidade temporal no teste e reteste, assim como índices satisfatórios de consistência interna e de análise de concordância entre avaliadores. Destaca-se, no entanto, que a atenção por si só pode e em geral flutua (Smith & Kosslyn, 2007).

Além disso, a concordância de juízes em relação às estratégias de cancelamento também se mostrou significativa. O coeficiente Kappa pode variar de 1 a -1, indicando concordância ou discordância completa, e o valor 0 indica um valor de concordância igual ao acaso. Para uma interpretação dos valores kappa, foram utilizadas as faixas de valores elaboradas por Landis e Koch (1977). Esses autores sugerem que os valores acima de 0,75, representam concordância excelente, valores abaixo de 0,40, uma concordância pobre e valores entre 0,40 e 0,75, uma concordância de suficiente a boa. Adicionalmente, no Capítulo 10, evidências quanto à validade por estrutura interna dos TCS-1 e TCS-2 são mostradas, acrescentando mais evidências de fidedignidade a ambas as versões do instrumento. Ressalta-se mais uma vez que o método de teste-reteste, além de suas desvantagens classicamente reportadas, para instrumentos de exame da atenção concentrada focalizada e, às vezes, atenção executiva (quando a tarefa é mais esforçosa), mostra-se especialmente limitado. Isso porque a atenção é um construto que pode flutuar e variar ao longo do tempo (Franzen, 2000).

REFERÊNCIAS

- Bickerton, W. L., Samson, D., Williamson, J., & Humphreys, G. W. (2011). Separating forms of Neglect using the Apples Test: validation and functional prediction in chronic and acute stroke. *Neuropsychology*, 10, 1-15.
- Chaves, M. L., & Izquierdo, Y. (1992). Differential diagnosis between dementia and depression: A study of efficiency increment. *Acta Neurologica Scandinavia*, 85, 378-382.
- Cunha, J. A. (2001). *Manual da versão em português das escalas Beck. Tradução e adaptação brasileira*. São Paulo: Casa do Psicólogo.
- Fonseca, R. P., Zimmermann, N., Pawlowski, J., Oliveira, C. R., Gindri, G., Scherer, L. C., Rodrigues, J. C., & Parente, M. A. M. P. (2012). Métodos em avaliação neuropsicológica: pressupostos gerais, neurocognitivos, neuropsicolinguísticos e psicométricos no uso e desenvolvimento de instrumentos. In J. Landeira-Fernandez, & S. Fukusima (Eds), *Métodos de pesquisa em neurociência clínica e experimental*. São Paulo: Manole.
- Franzen, M. D. (2000). *Reliability and validity in neuropsychological assessment*. New York: Kluwer.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis* (7th ed.). Harlow, UK: Pearson Education.

- Hedge, C., Powell, G., Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 1-21
- Landis, J., & Koch, G. G. (1997). The measurement of agreement for categorical data. *Biometrics*, 33(3), 159-179
- Lee, B. H., Kang, S. J., Park, J. M., Son, Y., Lee, K. H., Adair, J. C., Heilman, K. M., & Na, D. L. (2004). The Character-line bisection task: a new test for hemispatial neglect. *Neuropsychologia*, 42, 1715-1724.
- Manly, T., Dove, A., Blows, S., George, M., Noonan, M. P., Teasdale, T. W., ... Warburton, E. (2009). Assessment of unilateral spatial neglect: Scoring star cancellation performance from video recordings – method, reliability, benchmarks, and normative data. *Neuropsychology*, 23(4), 519-528. <http://doi.org/10.1037/a0015413>.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochem. Med.*, 22(3), 276-82.
- McHugh, R. K., & Barlow, D. H. (2012). The reach of evidence-based psychological interventions. In R. K. McHugh & D. H. Barlow (Eds.), *Dissemination and Implementation of Evidence-Based Psychological Interventions* (pp. 3-15). New York: Oxford University Press.
- Pasquali, L. (2009). Psicometria. *Revista da Escola de Enfermagem*, 43(Esp), 992-999.
- Pasquali, L. (2007). Validade dos testes psicológicos: será possível reencontrar o caminho? *Psicologia: Teoria e Pesquisa*, 23(Esp), 99-107.
- R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Revelle, W. (2015). *psych: Procedures for Personality and Psychological Research*. Evanston, Illinois. Retrieved from <http://cran.r-project.org/package=psych> Version = 1.5.8.
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega and the glb: comments on Sijtsma. *Psychometrika*, 74(1), 145-154.
- Smith, E. E., & Kosslyn, S. M. (2007). *Cognitive psychology: mind and brains*. Nova York: Pearson Education.
- Strauss, E., Sherman, E.M.S., Spreen, O. (2006). *A compendium of neuropsychological tests – administration, norms, and commentary* (3a ed.). New York: Oxford University Press.
- Urbina, S. (2014). *Essentials of psychological testing* (2nd ed.). Hoboken: Wiley.
- Urbina, S. (2007). *Fundamentos em testagem psicológica*. Porto Alegre: Artmed.
- Watkins, H. M. (2017). System justification in social psychology? A Survey of Responses to the Replicability Debate in Psychology. *PsyArXiv*. Retrieved from: <http://doi.org/10.17605/OSF.IO/NN67y>.
- Zanon, C. & Hauck Filho, N. (2015). Fidedignidade. In C. S. Hutz, D. R. Bandeira, & C.M., Trentini. *Psicometria* (pp. 85-95). Porto Alegre: Artmed.