
Revisão

Evidências de validade e fidedignidade em instrumentos de cancelamento

Evidence of validity and reliability in cancellation tests

Cristina Elizabeth Izabal Wong[✉], Charles Cotrena[✉] e Rochele Paz Fonseca[✉]

Grupo Neuropsicologia Clínica Experimental (GNCE), Programa de Pós-Graduação em Psicologia Ênfase em Cognição Humana, Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS), Porto Alegre, Rio Grande do Sul, Brasil

Resumo

Na presente revisão sistemática, objetivou-se apresentar um panorama dos estudos que buscaram evidências de validade e fidedignidade em instrumentos padronizados de cancelamento visual. Foram incluídos onze estudos com dados referentes a dez instrumentos de cancelamento visual. Destes, seis testes foram utilizados com objetivo principal de buscar evidências de validade e de fidedignidade. Observou-se que os testes de cancelamento mais utilizados para a obtenção de parâmetros psicométricos foram o *Apples Test*, *Character-Line Bisection Task* e *Symbol Cancellation Test*. Em relação aos tipos de validade mais explorados, foram a validade de construto e a validade concorrente/convergente. A técnica mais utilizada para sua obtenção foi a correlação com testes considerados padrão-ouro. Em relação à fidedignidade, a técnica mais utilizada foi a de teste-reteste. Em geral, observaram-se poucos estudos com o intuito de obter evidências de validade e de fidedignidade em instrumentos de cancelamento, sendo um número menor ainda em populações saudáveis. Estes parâmetros psicométricos e metodológicos precisam ser mais explorados para garantia de melhor acurácia dos instrumentos de avaliação neuropsicológica. © Cien. Cogn. 2012; Vol. 17 (2): 155-167.

Palavras-chave: validade; fidedignidade; testes de cancelamento, avaliação neuropsicológica.

Abstract

This systematic review aimed to provide an overview of studies that obtained evidence of validity and reliability of standardized visual cancellation instruments. We included eleven studies presenting data about ten visual cancellation tasks. six of these tests were target of a main goal of searching for evidence of their validity and reliability. The tests more explored regarding psychometric parameters were Apples Test, Character-Line Bisection Task and Symbol Cancellation Test. The most studied validity types were construct and concurrent convergent validities, mainly based on correlation with gold-standard tests. With regards to reliability, the most used technique was test-retest. In general there are few studies showing the interest in obtaining evidence of validity and reliability of cancellation instruments, with surprisingly even fewer evaluating healthy populations. These methodological and psychometric parameters need to be further used in order to guarantee better accuracy of neuropsychological assessment instruments. © Cien. Cogn. 2012; Vol. 17 (2): 155-167.

[✉] - **C.E.I. Wong** – Av. Ipiranga, 6681, n. 11, Sala 932, Partenon, Porto Alegre, RS 90.619-900, Brasil. E-mail para correspondência: cristina_izabal@yahoo.mx; **C. Cotrena** – E-mail para correspondência: charlescotrena@brturbo.com.br; **R.P. Fonseca** – E-mail para correspondência: rochele.fonseca@gmail.com.

Keywords: *validity; reliability; cancellation test, neuropsychological assessment.*

1 Introdução

A atenção é uma das funções que vem sendo mais exploradas, uma vez que as habilidades de busca visual desempenham um papel importante no dia-a-dia das pessoas e que envolvem a recepção, a todo o momento, de estímulos recebidos do ambiente (Cecilio-Fernandes & Rueda, 2007). Dentre os instrumentos de avaliação da atenção visual, os mais conhecidos e utilizados são os que envolvem o paradigma de cancelamento (Huang & Wang, 2008) em que a realização da tarefa requer que o participante realize um escaneamento visual. Geralmente as tarefas são compostas por dígitos, letras, figuras e símbolos. A tarefa do participante é discriminar os alvos dentre outras figuras com velocidade e desempenho acurado (Krinsky-McHale, Devenny & Kitter, 2008). As tarefas de cancelamento podem mensurar diversas funções cognitivas como a atenção concentrada, seletiva e dividida, a velocidade de processamento e estratégias de busca e habilidades práticas (Byrd, Touradji, Tang & Manly, 2004; Huang & Wang, 2009).

Tradicionalmente, os instrumentos de cancelamento são utilizados para avaliar possível heminegligência visual. Este quadro sindrômico pode envolver sintomas de atenção visual, negligência motora, sensorial e auditiva com possível prejuízo do comportamento (Azouvi *et al.*, 2002; Chiba, Yamaguchi & Eto, 2005; Vanier *et al.*, 1990). Entre suas características, apresenta heterogeneidade e geralmente ocorre devido a uma lesão cerebral unilateral, particularmente ao hemisfério direito, afetando o processamento das informações que chegam pelo campo sensorial contralateral esquerdo (Saj, Honore, Braem, Bernati & Rousseaux, 2012). Estudos com pacientes não acometidos por lesão vem sendo realizados ao explorar o desempenho em tarefas de cancelamento em quadros psiquiátricos, tais como bulimia nervosa (van den Eynde *et al.*, 2012), transtorno de déficit de atenção e hiperatividade (Hale *et al.*, 2009) e esquizofrenia (Liu, Chen, Yang, Chen & Lin, 2011). Na atualidade na área da neuropsicologia, vem crescendo o número de estudos acerca da obtenção de parâmetros de validade e fidedignidade de instrumentos de avaliação neuropsicológica com o intuito de que estes sejam válidos e fidedignos em relação aos componentes cognitivos a que se propõem avaliar. Desta forma, os procedimentos adotados devem obedecer aos parâmetros científicos adequados de obtenção de evidências de validade e fidedignidade na tentativa de mensurar com acurácia as funções cognitivas avaliadas (Peña-Casanova, Monllau & Gramunt, 2007). A importância das questões psicométricas, na avaliação, vem sendo estudada ainda em baterias de testes com adultos (Azouvi, Oliver, Montety, Samuel, Louis-Dreyfus & Tesio, 2003) e na adaptação de instrumentos em crianças (Argollo *et al.*, 2009). Portanto, com a utilização de instrumentos propriamente adaptados para uma população em específico, diminui a ocorrência de falsos-positivos no diagnóstico do paciente na prática clínica (Ostrosky-Solís, Ardila & Rosselli, 2006). Para que um instrumento possa ser considerado válido, este deve mensurar o componente a que se propõe avaliar (Paquali, 2009). Há diversos tipos de validade, as mais analisadas nos estudos são a validade de conteúdo, construto (convergente e discriminante) e critério. Conceitua-se a validade de conteúdo ao quão o instrumento analisa o que pretende medir; a validade de construto é caracterizada pela medida que o teste representa o construto avaliado com base na teoria; por fim, a validade de critério refere-se à medida que o instrumento é capaz de prever o desempenho quando comparado com grupos específicos (Pawlowski, Trentini & Bandeira, 2007). De igual importância psicométrica, a fidedignidade constitui-se da obtenção de evidências acerca da proporção em

que os escores estão livres de erros e como eles não variam com o tempo (Pasquali, 2007; Urbina, 2007). Neste contexto, ressalta-se a importância que os testes sejam devidamente adaptados em populações saudáveis quando são provenientes de outros países para que possam medir o construto pelo qual foram construídos (Pawlowski *et al.*, 2007).

Frente à importância dos instrumentos de cancelamento, no contexto da avaliação neuropsicológica, há a utilização de métodos que englobam o cuidado metodológico em processos de adaptação e construção de ferramentas de avaliação que seguem critérios científicos de obtenção de evidências de validade e de fidedignidade para diferentes populações. O objetivo do presente artigo, por meio de uma revisão sistemática, foi apresentar um panorama dos estudos que buscam evidências de fidedignidade e/ou de validade para instrumentos padronizados com tarefas de cancelamento visual. Para tanto, visa-se responder algumas questões de pesquisa:

- (1) quais são os instrumentos de cancelamento visual que tem sido mais utilizados em estudos com fins psicométricos?
- (2) quais são os tipos de parâmetros e as técnicas de análise mais utilizadas em estudos de validade e de fidedignidade com tais instrumentos? e
- (3) quais as características das amostras participantes destes estudos psicométricos?

2 Método

A revisão sistemática foi conduzida entre janeiro e março de 2012 nas bases de dados PubMed/MEDLINE e PsycINFO. Foram buscados *abstracts* de estudos empíricos conduzidos com instrumentos de avaliação da atenção por paradigma de cancelamento visual. Os estudos deveriam ter como objeto de investigação o estabelecimento de parâmetros psicométricos de instrumentos ou, como consequência de outros objetivos, apresentarem contribuições psicométricas ao processo de construção e/ou adaptação de testes de avaliação da atenção pelo paradigma de cancelamento visual. As palavras-chave utilizadas para as buscas foram: “cancellation task” *or* “cancellation test” *or* “visual attention” *and* “stability” *or* “reliability” *or* “precision” *or* “test retest” *or* “internal consistency” *or* “validity” *or* “validation”. Para a presente revisão sistemática foram analisados os resumos que foram publicados durante os anos de 2001 a 2012.

Inicialmente foram encontrados 905 resumos, dos quais 176 foram publicações repetidas entre bases. Para a seleção dos artigos que foram incluídos, fez-se análise duplo cego de cada *abstract*, por dois neuropsicólogos independentes, com consenso de um terceiro avaliador. Este procedimento obteve 95,5% de concordância pelo método simples (Andres & Marzo, 2004). Na análise de 22 *abstracts*, não houve consenso na inclusão de um, sendo posteriormente excluído pela análise do terceiro avaliador. Foram analisados, então, 21 textos completos, com ênfase nas seções Introdução, Método, Resultados e aspectos referentes às limitações na seção de Discussão, permanecendo 11 artigos para a presente revisão. Na Figura 1, é apresentado um fluxograma da análise sistemática de *abstracts* e artigos.

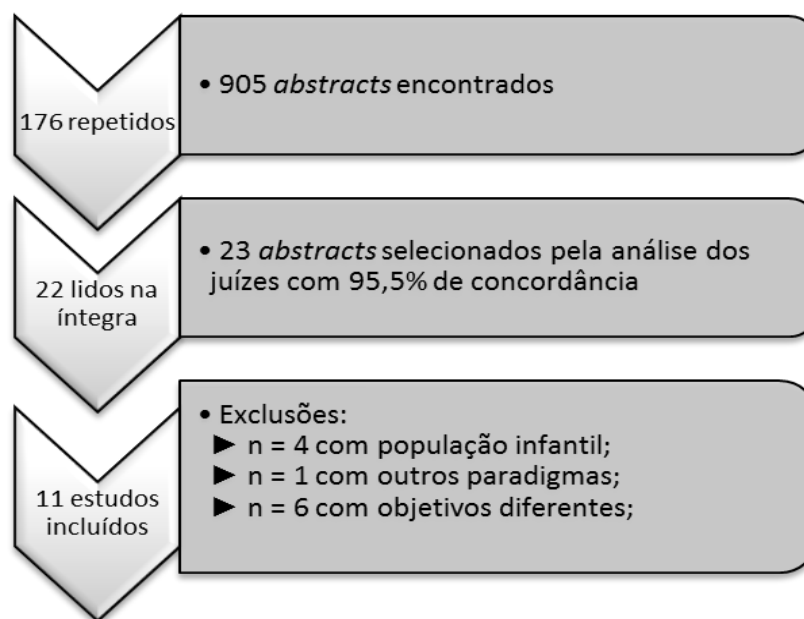


Figura 1 - Fluxograma de análise de *abstracts* e estudos completos incluídos na revisão sistemática.

3 Resultados

A partir dos objetivos da presente revisão sistemática, serão apresentados em um primeiro momento, os resultados em relação ao número de artigos encontrados por base de dados, além do número de *abstracts* nas diferentes combinações de palavras-chave utilizadas. Na Tabela 1 são apresentados os resultados para as buscas realizadas com as palavras-chave “cancellation test” and “stability” or “reliability” or “precision” or “test retest” or “internal consistency”, or “validity” or “validation”. Da mesma forma com “cancellation task” and, e “visual attention” and e todas as outras combinações de palavras.

Tabela 1 - Quantidade de *abstracts* encontrados por palavra-chave e base de dados. (Nota: Foram considerados todos os artigos, inclusive aqueles repetidos).

Palavras-chave	Bases de dados		
	PubMed	PsycINFO	Total
cancellation test and			
stability	2	0	2
reliability	5	1	6
precision	0	0	0
test retest	0	0	0
internal consistency	1	0	1
validity	8	2	10
validation	9	1	10
total parcial de artigos por base	25	4	29
cancellation task and			
stability	3	0	3
reliability	8	1	9

precisión	2	0	2
test retest	1	0	1
internal consistency	3	0	3
validity	15	1	16
validation	19	0	19
total parcial de artigos por base	51	2	53
<hr/>			
visual attention and			
stability	114	11	125
reliability	103	19	122
precision	94	4	98
test retest	45	3	48
internal consistency	20	7	27
validity	276	16	292
validation	95	16	111
total parcial de artigos por base	747	76	823
Total	823	82	905

Pode-se observar mediante a Tabela 1 que o maior número de estudos encontrados foi referente ao construto validade, sendo $n=318$ para “*validity*” e $n=140$ para “*validation*”. No entanto, as palavras com menos resumos encontrados foram “*internal consistency*”, com $n=31$, seguida por “*test retest*” com $n=49$. Além disso, na combinação de palavras-chave com “*cancellation test and*” encontrou-se o menor número de “*abstracts*”. De maneira geral, na base de dados PsycINFO obteve-se um escasso número de artigos com as diferentes combinações empregadas.

Todos os artigos incluídos foram publicados na língua inglesa. Nestes 11 estudos, identificou-se 10 diferentes paradigmas de cancelamento descritos brevemente a seguir, sendo os de número 1, 4, 5, 6, 7 e 9 objetos diretos de busca por evidências psicométricas e os de número 2, 3, 8 e 10 padrões-referência para análises comparativas ou correlacionais com os instrumentos testados.

1. *The Apples test* (Bickerton, Samson, Williamson & Humphreys, 2011). No teste é apresentada uma folha A4 com 150 maçãs, cuja distribuição é pseudoaleatória. Dois terços das maçãs são distratores (a metade com a maçã incompleta faltando uma parte do lado esquerdo e a outra metade faltando um pedaço da parte direita) o resto são os alvos (maçãs completas). O participante é instruído a cancelar todas as maçãs completas. No teste são considerados o número total de alvos cancelados
2. *The Bells test* (Gauthier *et al.*, 1989). O teste consiste numa folha A4, com 35 sinos (alvos) dentre 280 distratores distribuídos de forma pseudoaleatória por sete colunas. A tarefa do participante é cancelar todos os sinos que encontrar o mais rápido possível sem limite de tempo para a tarefa. Na pontuação são considerados o número total de omissões e a coluna do primeiro sino cancelado, além da estratégia de busca (organizada *versus* desorganizada).
3. *Line Bisection* (Azouvi *et al.*, 2006). É apresentada uma folha A4, com quatro linhas com diferentes medidas (5 cm, e 20 cm). O participante deve marcar um traço que divida cada linha em duas metades iguais. Neste teste os escores são medidos em mm,

- considerando-se desempenho positivo quando a distribuição está concentrada na direita e negativo, na esquerda.
4. *The Letter cancellation task* (Weintraub & Mesulam 1985). Este teste de lápis e papel é composto por 60 alvos (letras), que estão distribuídos na folha misturados com outras letras (distratores). Existem diferentes formas do teste e elas variam dependendo das instruções: por exemplo, busca por diferentes letras ou símbolos e/ou diferentes velocidades de escaneamento. Para a pontuação do teste, considera-se o tempo total da tarefa e o número total de omissões.
 5. *The Character-line bisection task (CLBT) – Letter-line and Star-line bisection* (Na, Adair, Choi, Seo, Kang & Heilman, 2000). Há duas seções de tarefas: na primeira o indivíduo deve cancelar o alvo mais próximo do centro da linha, inibindo os distratores. Os escores são quantificados por acertos.
 6. *The Computerized visual search test (CVST)*; Reid, Babani & Jon, 2009). O teste inclui seis tarefas que variam dependendo do tipo de distrator utilizado (alto e baixo) e do número de objetivos (busca de memória). Cada tarefa inclui 30 estímulos na tela do computador. O teste tem dois níveis e três tarefas, cada uma com diferentes distratores, alguns mais relacionados do que outros. Na tela são mostradas oito embalagens de sopa (vermelhas e brancas) e 22 distratores que são da cor contrária à do alvo. A pessoa tem de procurar um tipo específico de embalagem dentre as outras. Conforme o aumento da dificuldade da tarefa, o participante tem que procurar dois ou três tipos diferentes de sopa. O teste é medido pelo número de erros cancelados (distratores).
 7. *D2* (Brickenkamp & Zillmer, 1998). É uma tarefa de lápis e papel, com 14 linhas, cada uma com 47 letras “p” e “d”. As letras podem ter uma ou duas aspas, na sua parte superior ou inferior. O participante tem de cancelar todos os “d”. Os escores são quantificados pelo número total de acertos e erros do teste.
 8. *Star Cancellation test* (Wilson, Cockburn & Halligan, 1988). O teste consiste em uma folha com estrelas distribuídas. A tarefa do participante na versão modificada é cancelar 56 estrelas dentre outras figuras como triângulos isósceles em cor preta. Na pontuação do teste são considerados a primeira estrela cancelada, estratégia utilizada, número total de alvos cancelados e o tempo da prova.
 9. *Symbol Cancellation test (SCT)* (Solfrizzi *et al.*, 2002). O teste contém figuras geométricas (quadrados), com oito categorias diferentes discriminadas pela posição de uma linha interna, por exemplo, dentro de alguns quadrados há uma linha inclinada no canto superior direito. O participante tem que cancelar os alvos que são apresentados no topo da tela. Os escores são medidos pelo número de acertos.
 10. *Zazzo’s Cancellation test (ZCT)* (Zazzo, 1974). Nesta versão de oito linhas do teste, avalia-se a habilidade do participante para riscar todos os 29 alvos (signos), dentre 200 distratores. O examinando tem de cancelar os alvos o mais rápido possível. Para a análise do teste é considerado o tempo da prova.

São apresentados, na Tabela 2, os estudos que apresentam objetivo de buscar evidências de validade. Os estudos foram reunidos por paradigma de cancelamento em comum.

Conforme os resultados mostrados na Tabela 2, foram oito estudos que analisaram no mínimo um tipo de validade. Todos os estudos com investigação psicométrica focaram-se na busca de evidências de validade de construto, seja concorrente/convergente ou discriminante/divergente. A técnica utilizada por todos, exceto por Reid, Babani e Jon (2009), foi de correlação com desempenho mensurado por outros testes que avaliam o mesmo construto ou construto oposto. Em relação aos participantes somente dois estudos Bates e Lemay (2004) e Uttil e

Pilkenton-Taylor (2001) incluíram adultos saudáveis, analisando os outros o desempenho de populações clínicas, controlado pela performance de controles saudáveis. Todos os instrumentos foram na modalidade lápis e papel, havendo apenas um estudo em que desenvolveram uma tarefa informatizada de cancelamento (Reid *et al.*, 2009). Apresentam-se na Tabela 3 os estudos com objetivo de verificar indícios de fidedignidade do instrumento.

Paradigma de Cancelamento	Autores	Objetivos	Tipo de validade/Técnica utilizada	Participantes	Resultados
Apples Test	Bickerton <i>et al.</i> (2011)	Obter evidências de validade.	Validade concorrente/Correlação com outro teste que avalia o mesmo construto, Kruskal-Wallis test.	N=86 controles, idades 47-88 anos, N=7 LHE, N=18 LHD, idades 36-74 anos.	Encontrou-se correlações significativas entre os testes.
Bells Test	Azouvi, <i>et al.</i> (2006)	Adaptar e obter validade da bateria de testes (CBS).	Validade concorrente/Correlações inter-teste, análise entre a CBS e outros testes. Consistência interna, Spearman.	N=576 adultos saudáveis de 20 anos, N=206 LHD e N=89 LHE.	A bateria mostrou-se válida para avaliação de heminegligência.
	Rorden <i>et al.</i> (2010)	Desenvolver software (CoC) para a interpretação dos escores.	Correlação entre os escores do CoC e outros testes utilizados.	N=110 pacientes com LHD.	Com o software houve maior acurácia diagnóstica em 98% dos casos.
Character-line Bisection Task (CLBT)	Lee <i>et al.</i> (2004)	Demonstrar a validade do teste.	Validade concorrente/Correlações do CLBT com outros testes. Correlação de Pearson.	N=81 controles, idade 42-81 anos, N=80 pacientes idade 32-82 anos.	O CLTB mostrou altas correlações com testes que avaliam o mesmo construto.
Computerized Visual Search Test (CVST)	Reid <i>et al.</i> (2009)	Descrever o teste e obter validade com o estudo piloto.	Comparação com outro teste de lápis e papel. Correlação de Spearman.	N=20 adultos saudáveis (M=25,5) e N=5 pacientes com lesão cerebral (M=26,6).	O CVST é sensível na detecção de déficits atencionais.
D2 Test	Bates <i>et al.</i> (2004)	Examinar a validade de construto e validade interna.	Validade de construto, convergente e discriminante/Correlações com outros testes que avaliam os mesmos construtos. Consistência interna, Cronbach's alpha.	N=364 adultos saudáveis, idade 28-32 anos.	O D2 mostrou-se válido na avaliação da atenção visual e da velocidade.
Letter Cancellation Task	Rorden <i>et al.</i> (2010)	Desenvolver um software (CoC) para a interpretação dos escores.	Correlação entre os escores do CoC e outros testes utilizados.	N=110 pacientes com LHD.	Com o software houve maior acurácia diagnóstica em 98% dos casos.
Line Bisection	Uttl <i>et al.</i> (2001)	Criar normas para o teste e mostrar evidências de validade.	Validade convergente e divergente/Correlações com outros testes. Correlação Spearman.	N=351 adultos saudáveis, idade 18-91 anos.	O teste mostrou-se válido.
	Azouvi, <i>et al.</i> (2006)	Desenvolver uma bateria de testes e obter parâmetros a validade.	Validade concorrente/Correlações inter-teste, da CBS com outros testes. Consistência interna, Spearman.	N=457 adultos saudáveis de 20 anos, N=206 LHD e N=89 LHE.	A bateria mostrou-se válida na avaliação da heminegligência.
Star Cancellation Test	Woods <i>et al.</i> (2007)	Quantificar as estratégias de cancelamento do teste a fim de avaliar a validade.	Validade convergente/Correlação com os escores do teste. Correlação de Spearman.	N=21 controles, idade 42-81 anos, N=21 LHD e N=8 LHE, idades 38-84 anos.	Verificou-se parâmetros de validade com base nas estratégias utilizadas.
Symbol Cancellation Test (SCT)	Solfrizzi <i>et al.</i> (2002)	Avaliar a validade convergente e discriminante do SCT.	Validade convergente e discriminante/Análises dos escores em relação a idade, escolaridade e sexo. Correlações com outros instrumentos.	N=232 controles idosos 68-87 anos e N=34 pacientes com DA.	O teste é válido e pode ser utilizado em populações com DA.

Tabela 2 - Estudos que buscam evidências de validade por instrumento de cancelamento. Nota: DA= Demência do tipo Alzheimer, LHD= Lesão de hemisfério direito, LHE= Lesão de hemisfério esquerdo, CBS= *Catherine Bergego Scale*, CoC= *Center of Cancellation*.

Paradigma de Cancelamento	Autores	Objetivos	Técnica utilizada	Participantes	Resultados
Apples Test	Bickerton <i>et al.</i> (2011)	Obter parâmetros de fidedignidade.	Teste-reteste. Correlação entre as duas aplicações.	N=20 pacientes, idade 36-74 anos.	O teste é fidedigno para a avaliação de pacientes com heminegligência.
Character-line Bisection Task (CLBT)	Lee <i>et al.</i> (2004)	Analisar a fidedignidade do instrumento.	Teste-reteste. Intervalo entre as aplicações de 24 horas. Correlação com o coeficiente de Pearson's.	N=80 Pacientes, idade 32-84 anos, N=81 controles idade 42-81 anos.	CLBT mostrou-se fidedigno por meio da análise do teste-reteste.
Letter Cancellation Test	Uttl <i>et al.</i> (2001)	Obter normas para o teste e mostrar evidências de fidedignidade.	Correlação de Spearman mediante duas formas paralelas do teste.	N=351 adultos saudáveis, idade 18-91 anos.	Obteve-se evidências de fidedignidade frente a correlações significativas.
Star Cancellation	Manly <i>et al.</i> (2009)	Buscar evidências de fidedignidade, por meio de gravações em vídeo.	Correlação entre os escores obtidos no momento da avaliação e gravação realizada.	N=18 pacientes com heminegligência e N=19 grupo controle.	O instrumento mostrou-se fidedigno e sensibilidade do instrumento.
Symboll Cancellation Test (SCT)	Solfrizzi <i>et al.</i> (2002)	Analisar a fidedignidade do instrumento.	Correlação com outros testes.	N=34 pacientes com DA e N=232 controles idosos 68-87 anos, pareados por idade, escolaridade e sexo.	Mostrou-se a fidedignidade do instrumento.
Zazzo's Cancellation Test (ZCT)	Fabrigoule <i>et al.</i> (2003)	Obter a fidedignidade entre avaliadores psicólogos e médicos clínicos gerais.	Correlações entre escores proporcionados pelos os avaliadores.	N=35 pacientes idosos (M=73; DP=9,4).	Encontrou-se correlação entre os diagnósticos oferecidos por ambos os avaliadores.

Tabela 3 - Estudos que exploraram parâmetros de fidedignidade de instrumentos de cancelamento.

Na avaliação da fidedignidade, a técnica mais utilizada foi do teste-reteste. Estes estudos incluíram como participantes principalmente grupos clínicos (Bickerton *et al.*, 2011; Fabrigoule, Lechevallier, Crasborn, Darthigues & Orgogozo, 2003; Lee *et al.*, 2004; Manly *et al.*, 2009; Solfrizzi *et al.*, 2002). Por fim, dos 11 estudos, cinco preocuparam-se apenas com evidências de validade, dois com indícios de fidedignidade e quatro com ambos os parâmetros. Dos 10 instrumentos de avaliação por cancelamento, 3 obtiveram evidências apenas de validade, 1 com evidências de fidedignidade e apenas em 5 instrumentos foram apresentados dados referentes a evidências de validade e fidedignidade dos mesmos.

4 Discussão

O presente estudo de revisão sistemática buscou por investigações que utilizassem parâmetros psicométricos em estudos com instrumentos de cancelamento visual. De acordo com a primeira questão que norteou esta revisão sistemática, (1) quais são os instrumentos de cancelamento visual que tem sido mais utilizados com fins psicométricos? Observou-se que os instrumentos de cancelamento mais frequentemente utilizados são, em ordem decrescente, o *Star Cancellation Test* (Bickerton *et al.*, 2011; Lee *et al.*, 2004; Manly *et al.*, 2009; Woods *et al.*, 2007) o *Bells Test* (Azouvi *et al.*, 2006; Rorden *et al.*, 2010) e o *Letter Cancellation Task* (Rorden *et al.*, 2010; Uttl *et al.*, 2001). Para realização de estudos psicométricos os mais utilizados foram o *Apples Test* (Bickerton *et al.*, 2011), o CLBT (Lee *et al.*, 2004), D2 (Bates *et al.*, 2004), *Letter cancellation task* (Uttl *et al.*, 2001) e SCT (Solfrizzi *et al.*, 2002). Por meio da observação dos achados em sua totalidade, pôde-se perceber que há poucos instrumentos que contam com procedimentos de obtenção de evidências de validade e fidedignidade para as diferentes populações para cuja avaliação são desenvolvidos. Tal fato pode comprometer as capacidades diagnósticas das ferramentas de avaliação assim como a

melhor compreensão dos déficits por elas mensurados (Chiang & Gau 2008; Huang *et al.*, 2009).

Mais especificamente, nesta revisão pode-se observar que muitos dos estudos utilizaram outros testes como padrão-ouro para obter a validade do teste alvo. O mais utilizado foi o *Star Cancellation Test* (Bickerton *et al.*, 2011; Lee *et al.*, 2004; Manly *et al.*, 2009; Woods *et al.*, 2007), seguido pelo *Bells Test* (Azouvi *et al.*, 2006; Rorden *et al.*, 2010). Estes testes são comumente utilizados na avaliação da atenção, na medida em que tem se mostrado muito sensíveis na detecção de déficits atencionais e da síndrome de heminegligência (Azouvi *et al.*, 2002; Bailey, Riddoch & Crome, 2005; Karnath, Baier & Nägele, 2005; Vanier *et al.*, 1990). Assim, dos 10 instrumentos utilizados nos estudos, pode-se perceber que houve somente seis instrumentos que foram alvos para a procura de características psicométricas dos testes. Os demais quatro instrumentos foram utilizados como parâmetros de comparação com os testes utilizados (Azouvi *et al.*, 2006; Bickerton *et al.*, 2011; Lee *et al.*, 2004; Manly *et al.*, 2009; Rorden *et al.*, 2010), com o intuito de auxiliar na interpretação dos escores no desenvolvimento de um software (Rorden *et al.*, 2010), na comparação com uma tarefa no computador (Reid *et al.*, 2009) e, por último, um deles foi utilizado para verificar a concordância entre avaliadores (Fabrigoule *et al.*, 2003). Ressalta-se, ainda, que de todos os estudos analisados encontrou-se um teste de cancelamento no qual a tarefa foi realizada no computador. Nesta investigação, foram adotados os mesmos parâmetros utilizados em estudos que envolveram os instrumentos de cancelamento, sendo uma alternativa aos tradicionais testes de lápis e papel demonstrando as características psicométricas do teste (Reid *et al.*, 2009).

Em relação à questão (2) quais são os tipos de parâmetros e as técnicas de análise mais utilizadas em estudos de validade e de fidedignidade com tais instrumentos? No que diz respeito à validade dos testes, o tipo mais analisado foi a validade de construto, mais especificamente a validade concorrente/convergente. Com referência às técnicas utilizadas para a obtenção de validade, a mais utilizada foi a correlação com o desempenho de testes considerados padrão-ouro, que avaliam o mesmo construto (Azouvi *et al.*, 2006; Bates *et al.*, 2004; Bickerton *et al.*, 2011; Fabrigoule *et al.*, 2003; Lee *et al.*, 2004; Manly *et al.*, 2009; Reid *et al.*, 2009; Rorden *et al.*, 2010; Woods *et al.*, 2007). No entanto, ainda que alguns destes estudos realizassem comparações com amostras clínicas e grupos controle, poucas explicitaram sobre a validade de critério do teste. A comparação de amostras saudáveis com populações clínicas é essencial para a neuropsicologia, uma vez que tais tarefas são desenvolvidas para diagnósticos da função atencional contribuindo para a detecção de heminegligência em pacientes neurológicos.

Em continuidade, dando resposta à questão acima mencionada em relação à busca de evidências de fidedignidade, a técnica mais utilizada foi a de teste-reteste (Bickerton *et al.*, 2011; Lee *et al.*, 2004) e foram realizadas análises de correlação entre as pontuações dadas pelos avaliadores (Fabrigoule *et al.*, 2003; Manly *et al.*, 2009), os quais apontam a importância no cuidado da interpretação dos resultados dos instrumentos. Em relação à técnica mais utilizada nos estudos de teste-reteste nas tarefas de cancelamento, aponta-se que as características destas tarefas são sua constituição unitária e não sua mensuração por itens, o que dificulta em alguns dos casos em realizar análise mediante a comparação das duas metades ou formas paralelas do teste (Pasquali, 2003). Dos estudos que procuraram evidências de fidedignidade, apenas um fez comparação com duas formas paralelas do teste (Uttl *et al.*, 2001). Assim, por meio dos estudos mostra-se a necessidade de um maior controle no processo de avaliação e do uso das tecnologias com o intuito de aprimorar a acurácia nas avaliações (Manly *et al.*, 2009; Reid *et al.*, 2009; Solfrizzi *et al.*, 2002) e sobre a pontuação

dos mesmos mediante o uso de um software para novas formas de pontuação (Rorden *et al.*, 2010).

Por último, quanto à terceira questão, (3) quais as características das amostras participantes destes estudos psicométricos? Foram encontrados mais estudos realizados com populações clínicas, com grupos controles de adultos saudáveis, como em pacientes com LHD e LHE (Azouvi *et al.*, 2006; Reid *et al.*, 2009; Rorden *et al.*, 2010; Woods & Mark, 2007), com síndrome de heminegligência (Manly *et al.*, 2009), também pacientes com DA (Solfrizzi *et al.*, 2002) e com declínio cognitivo leve (Fabrigoule *et al.*, 2003). Nestes estudos, mesmo com a realização de comparações entre os grupos clínicos e grupos saudáveis, nenhum deles mostrou resultados específicos referentes à validade de critério. A validade de critério refere-se ao tipo de validade que prediz um critério externo (Pasquali, 2007). Portanto, na comparação dos grupos clínicos com controles, os dados poderiam ter sido enriquecidos com os estudos anteriormente apresentados. Em relação às características dos participantes, ainda, também houve estudos apresentando normas do desempenho do teste para populações de adultos saudáveis (Bates & Lemay, 2004; Lee *et al.*, 2004; Uttl *et al.*, 2001).

Há um considerável número de estudos realizados com instrumentos de cancelamento, porém das investigações realizadas poucas avaliam o desempenho em populações saudáveis (Huang & Wang, 2008). Diante disto, existe uma necessidade de verificar quais seriam as diferenças entre os grupos saudáveis e clínicos para um melhor entendimento dos processos envolvidos e, por conseguinte, auxiliar no diagnóstico e plano de tratamento adequado (Fabrigoule *et al.*, 2003; Fonseca, Salles & Parente, 2008). Conforme os artigos analisados, pode-se observar que houve estudos que encontraram diferenças significativas quanto as variáveis sociodemográficas como idade e escolaridade (Azouvi *et al.*, 2006; Bates *et al.*, 2004), fornecendo também evidências de validade de critério não explicitadas. Assim mais estudos de adaptação de instrumentos precisam ser realizados com normas observando variáveis sociodemográficas com o intuito de obter dados que melhor representem as populações a que se destinar avaliar. Assinala-se, neste contexto, a exploração de pontos de corte de acordo com a sensibilidade e especificidade de cada instrumento, procurando suprir uma importante lacuna ainda existente de dados psicométricos associados a modelos teóricos de dissociação e modularidade para interpretação neurocognitiva no contexto da prática em neuropsicologia clínica. Uma das limitações do estudo é referente aos tópicos discutidos sobre a validade, já que tem-se desenvolvido novas diretrizes sobre os tipos de validade referentes a qualidade dos critérios de reconhecimento e avaliação (Conselho Federal de Psicologia, 2011). Estes critérios, por sua recente existência, apenas estão sendo consideradas em novos estudos e não foram encontrados na presente revisão.

Referências bibliográficas

- Andres, A.M.; Marzo, P.F. (2004). Delta: a new measure of agreement between two raters. *Br. J. Math. Statistical Psychol.*, 57 (1), 1-19.
- Argollo, N., Bueno, O. F. A., Shayer, B., Godinho, K., Abreu, K., Durán, P., Assis, A., Lima, F., Silva, T., Guimaraes, J., Carvalho, R., Moura, I. & Seabra, A. G. (2009). Adaptação transcultural da Bateria NEPSY – Avaliação neuropsicológica do desenvolvimento: estudo-piloto. *Avaliação Psicológica*, 8 (1), 59-75.
- Azouvi, P.; Bartolomeo, P.; Beis, J-M.; Perennou, D.; Pradat-Diehl, P.; Rousseaux, M. (2006). A battery of tests for the quantitative assessment of unilateral neglect. *Restorative Neurol. Neurosci.*, 24, 273-285.

- Azouvi, P.; Oliver, S.; de Montety, G.; Samuel, S.T.; Louis-Dreyfus, A.; Tesio, L. (2003). Behavioral assessment of unilateral neglect: study of the psychometric properties of the Catherine Bergego Scale. *Arch. Phys. Med. Rehabil.*, 84, 51-57.
- Azouvi, P.; Samuel, C.; Louis-Dreyfus, A.; Bernati, T.; Bartolomeo, P.; Beis, J-M.; Chokron, S.; Leclercq, M.; Marchal, F.; Martin, Y.; de Montety, G.; Olivier, S.; Perennou, D.; Pradat-Diehl, P.; Prairial, C.; Rode, G.; Siéroff, E.; Wiart, L.; Rousseaux, M. (2002). Sensitivity of clinical and behavioural tests of spatial neglect after right hemisphere stroke. *J. Neurol. Neurosurg. Psychiatry*, 73, 160-166.
- Bailey, M.J.; Riddoch, M.J.; Crome, P. (2005). Test-retest stability of three tests for unilateral visual neglect in patients with stroke: Star Cancellation, Line Bisection, and Baking Tray Task. *Neuropsychological Rehabil.*, 14(4), 403-419.
- Bates, M.E.; Lemay, E. P. (2004). The D2 test of attention: construct validity and extension in scoring techniques. *J. Int. Neuropsychological Soc.*, 10, 392-400.
- Brickenkamp, R.; Zillmer, E. (1998). *The d2 test of attention*. Seattle, Washington: Hogrefe & Huber Publishers.
- 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*, 25 (5), 567-580.
- Byrd, D.A.; Touradji, P.; Tang, M.-X.; Manly, J.J. (2004). Cancellation test performance in African American, Hispanic, and White elderly. *J. Int. Neuropsychological Soc.*, 10, 401-411.
- Cecilio-Fernandes, D.; Rueda, F.J.M. (2007). Evidência de validade concorrente para o Teste de Atenção Concentrada (TEACO-FF). *Rev. Psicologia Vetor Editora*, 8(2), 167-174.
- Chiang, M.; Gau, S.S.F. (2008). Validation of attention-deficit-hyperactivity disorder subtypes among Taiwanese children using neuropsychological functioning. *Australian New Zealand J. Psychiatry*, 24, 526-535.
- Chiba, Y.; Yamaguchi, A.; Eto, F. (2005). A simple method to dissociate sensory-attentional and motor-intentional biases in unilateral visual neglect. *Brain Cogn.*, 58, 269-273.
- Conselho Federal de Psicologia (2011). *Ano da Avaliação Psicológica, Textos Geradores*. Brasília D.F.
- Gauthier, L.; Dehaut, F.; Joanette, Y. (1989). The Bells test: a quantitative and qualitative test for visual neglect. *Int. J. Clin. Neuropsychol.*, 11(2), 49-54.
- Fabrigoule, C.; Lechevallier, N.; Crasborn, L.; Darthigues, J-F.; Orgogozo, J-M. (2003). Inter-rater reliability of scales and tests used to measure mild cognitive impairment by general practitioners and psychologists. *Cur. Med. Res. Opinion*, 19(7), 603-608.
- Fonseca, R.P.; Salles, J.F.; Parente, M.A.M.P. (2008). Development and content validity of the Brazilian brief neuropsychological assessment battery NEUPSILIN. *Psychol. Neurosci.*, 1(1), 55-62.
- Hale, J.B.; Reddy, L.A.; Decker, S.L.; Thompson, R.; Henzel, J.; Teodori, A.; Forrest, E.; Eusebio, E.; Denckla, B. (2009). Development and validation of an attention-deficit/hyperactivity disorder (ADHD) executive function and behavior rating screening battery. *J. Clin. Exp. Neuropsychol.*, 31 (8), 897-912.
- Huang, H-C.; Wang, T-Y. (2009). Stimulus edects on cancellation task performance in children with and without dyslexia. *Behav. Res. Meth.*, 41(2), 539-545.
- Huang, H-C.; Wang, T-Y. (2008). Visualized representation of visual search patterns for a visuospatial attention test. *Behav. Res. Meth.*, 40(2), 383-390.
- Karnath, H-O., Baier, B.; Nägele, T. (2005). Awareness of the functioning of one's own limbs mediated by the insular cortex? *J. Neurosci.*, 25(31), 7134-7138.

- Krinsky-McHale, S.J.; Devenny, D.A.; Kitter, P. (2008). Selective attention deficits associated with mild cognitive impairment and early stage Alzheimer's disease in adults with Down syndrome. *Am. J. Mental Retardation*, 115(5), 369-386.
- Lee, B.H.; Kang, S.J.; Park, J.M.; 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.
- Liu, Y-C. ; Chen, K-C.; Yang, Y.K. ; Chen, Y-L. ; Lin, K-C. (2011). Relationship between hemispatial inattention and performance of activities of daily living in patients with schizophrenia. *Perceptual Motor Skills*, 112 (3), 703-710.
- Manly, T.; Dove, A.; Blows, S.; George, M.; Noonan, M.A.P.; Dodds, C.M.; Fish, J.; Teasdale, T.W.; Warburton, E. (2009). Assessment of unilateral spatial neglect: scoring Star cancellation performance from video recordings-method, reliability, benefits, and normative data. *Neuropsychology*, 23(4), 519-528.
- Na, D.L. ; Adair, J.C.; Choi, S.H.; Seo, D.W.; Kang, Y.; Heilman, K.M. (2000). Ipsilesional versus contralesional neglect depends on attentional demands. *Córtex*, 36, 455-467.
- Ostrosky-Solís, F.; Ardila, A.; Rosselli, M. (2006). Test Review. *Arch. Clin. Neuropsychol.*, 21, 195-197.
- Pasquali, L. (2009). *Psicometria. Rev Esc Enferm*, 43(Esp), 992-999.
- Pasquali, L. (2007). Validade dos testes psicológicos: será possível reencontrar o caminho? *Psicologia Teoria Pesq.*, 23(Esp), 99-107.
- Pasquali, L. (2003). *Teoria dos testes na psicologia e na educação*. Vozes: Petrópolis.
- Pawlowski, J.; Trentini, C.M.; Bandeira, D.R. (2007). Discutindo procedimentos psicométricos a partir da análise de um instrumento de avaliação neuropsicológica breve. *Psico-USF*, 12(2), 211-219.
- Peña- Casanova, J.; Monllau, A.; Gramunt, N.F. (2007). La psicometría de las demencias a debate. *Neurología*, 22(5), 301-311.
- Reid, D.; Babani, H.; Jon, E. (2009). Development of a computerized visual search test. *Int. J. Rehabil. Res.*, 32(3), 205-212.
- Rorden, C.; Karnath, H-O. (2010). A simple measure of neglect severity. *Neuropsychologia*, 48, 2758-2763.
- Saj, A.; Honore, J.; Braem, B.; Bernati, T.; Rousseaux, M. (2012). Time since stroke influences the impact of hemianopia and spatial neglect on visual-spatial tasks. *Neuropsychology*, 26 (1), 37-44.
- Solfrizzi, V.; Panza, F.; Torres, F.; Capurso, C.; Díntrono, A.; Colacicco, A.M.; Carpurso, A. (2002). Selective attention skills in differentiating between Alzheimer's disease and normal aging. *J. Geriatric Psychiatry Neurol.*, 15, 99-109.
- Urbina, S. (2007). *Fundamentos em Testagem Psicológica*. Porto Alegre: Artmed.
- Uttl, B.; Pilkenton-Taylor, C. (2001). Letter cancellation performance across the adult life span. *Clin. Neuropsychologist*, 15(4), 521-530.
- Van den Eynde, F.; Samarawickrema, N.; Kenyon, M.; DeJong, H.; Lavender, A.; Startup, H.; Schmidt, U. (2012). A study of neurocognition in bulimia nervosa and eating disorder not otherwise specified-bulimia type. *J. Clin. Exp. Neuropsychol.*, 34 (1), 67-77.
- Vanier, M.; Gauthier, L.; Lambert, J.; Pepin, E. P.; Robillard, A.; Dubouloz, C.J.; Gagnon, R.; Yoannette, Y. (1990). Evaluation of left visuospatial neglect: norms and discrimination Power of two tests. *Neuropsychology*, 4, 87-96.
- Weintraub, S.; Mesulam, M.M. (1985). Mental state assessment of young and elderly adults in behavioral neurology. Em: Mesulam, M.M. (Ed.). *Principles of behavioral neurology*, (71-123). Philadelphia: Davis Company.

- Wilson, B.A.; Cockburn, J.; Halligan, P. (1988). *Behavioral Inattention Test*. Bury St. Edmunds, England: Thames Valley Test Company.
- Woods, A.J.; Mark, V.W. (2007). Convergent validity of executive organization measures on cancellation. *J. Clin. Exp. Neuropsychol.*, 29 (7), 719-723.
- Zazzo, R. (1974). *Test des deux barrages*. *Actualités Pédagogiques et Psychologiques*, Vol. 7. Neufchatel: Delachaux et Nestlé.