

Reliability and Construct Validity of the Bells Test

Cristina Elizabeth Izábal Wong¹

Universidad Autónoma de Sinaloa, Culiacán, México

Laura Damiani Branco, Charles Cotrena

Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brazil

Yves Joannette

University of Montreal, Montreal, Canada

Rochele Paz Fonseca

Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brazil

ABSTRACT

The Bells Test (BT) is widely used to aid in the diagnosis of heminegligence. The objective of this study was to evaluate the convergent validity of the BT, comparing it with tools that evaluate similar constructs, and to investigate its test-retest reliability. The sample included 66 healthy adults age 19-75 years. The reliability was evaluated through a test-retest procedure, with correlations and t-tests for paired samples, while validity was investigated through comparisons between the performance on the BT and scores in the Concentrated Attention test (CA-15), the Sustained Attention test (SA), and the WAIS-III Symbols and Codes subtests. Positive correlations were found between test and retest in both BT versions, as well as between the number of BT omissions and other attention measures. These results corroborate the validity and reliability of the two BT versions in the Brazilian population.

Keywords: attention; neuropsychological assessment; reliability; validity.

RESUMO – Fidedignidade e Validade de Construto do Teste de Cancelamento dos Sinos

O Teste de Cancelamento dos Sinos (TCS) é amplamente utilizado para auxiliar o diagnóstico de heminegligência. O objetivo deste estudo foi avaliar a validade convergente do TCS, comparando-o a ferramentas que avaliam construtos similares, e investigar sua fidedignidade teste-reteste. A amostra incluiu 66 adultos saudáveis com idades entre 19 e 75 anos. A fidedignidade foi avaliada por meio de procedimento teste-reteste, com correlações e testes *t* para amostras pareadas, enquanto a validade foi investigada através de comparações entre o desempenho no TCS e escores no teste de Atenção Concentrada (AC-15), teste de Atenção Sustentada (AS) e os subtestes Símbolos e Códigos do WAIS-III. Correlações positivas foram encontradas entre teste e reteste nas duas versões do TCS, assim como entre o número de omissões no TCS e demais medidas de atenção. Esses resultados corroboram a validade e fidedignidade das duas versões do TCS na população Brasileira.

Palavras-chave: atenção; avaliação neuropsicológica; fidedignidade; validade.

RESUMEN – Confiabilidad y Validez de Constructo del Test de Cancelación de las Campanas

El Test de las Campanas (TC) es un instrumento ampliamente utilizado para auxiliar en el diagnóstico de heminegligencia. El objetivo de este estudio fue evaluar la validez convergente del TCS, comparándolo con otras herramientas que evalúan constructos similares, e investigar su confiabilidad test-retest. La muestra incluyó 66 adultos con buena salud, de 19 a 75 años. La confiabilidad fue evaluada a través de procedimientos de test-retest, con correlaciones y tests-*t* para muestras pareadas, y la validez fue investigada a través de comparaciones entre el desempeño del TCS y los resultados en el test de Atención Concentrada (AC-15), test de Atención Sostenida (AS) y los sub-tests Símbolos y Códigos del WAIS-III. Correlaciones positivas fueron encontradas entre test y retest en las dos versiones del TCS así como entre el número de omisiones en el TCS y otras medidas de atención. Estos resultados corroboran validez y confiabilidad de las dos versiones del TCS en la población brasileña.

Palabras clave: atención; evaluación neuropsicológica; confiabilidad; validez.

Cancellation tasks are among the most commonly used techniques to detect visuospatial neglect (Bickerton, Samson, Williamson, & Humphreys, 2011;

Rorden & Karnath, 2010). This condition may result from right hemisphere lesions or other types of unilateral dysfunctions, and can have a significant impact on

¹ Endereço para correspondência: Grupo de Investigación en Procesos Básicos (GIPB), Facultad de Psicología. Universidad Autónoma de Sinaloa (UAS), Calzada de las Américas y Boulevard Universitarios s/n, Culiacán, Sinaloa. E-mail: cristina.izabalwong@gmail.com

Acknowledgments: The authors would like to thank Dr. Simone Barreto and Dr. Karin Zazo Ortiz for their contributions to this project.

Financial support: The authors received financial support from the Coordination for the Improvement of Higher Education Personnel (CAPES) and the National Council for Scientific and Technological Development (CNPq).

daily functioning. Regardless of its etiology, visuospatial neglect is usually characterized by a failure to respond to stimuli presented in the visual field contralateral to the lesion (Azouvi et al., 2006; Buxbaum, Dawson, & Linsley, 2012; Toglia & Cermak, 2007). Visuospatial neglect may lead patients to ignore targets on one half of a cancellation array, or to draw or copy only half the features of an image (Lee et al., 2004).

Cancellation tasks require both visual search and attentional engagement, allowing for the assessment of both selective and focused attention. As such, these tasks may help detect alterations in attention, perception and/or praxis (Alqahtani, 2015; Lee et al., 2004; Solfrizzi et al., 2002). The number of targets omitted in cancellation arrays has also proved to be a reliable and valid measure of the severity of visuospatial neglect (Toglia & Cermak, 2007). The visual search strategies used in cancellation tasks can also be used to evaluate executive functions, since visuospatial neglect does not have an impact on visual search per se, and disorganized strategies are generally the result of an underlying executive dysfunction (Woods & Mark, 2007).

The easy and quick administration of cancellation tasks, combined with the simplicity of their instructions, facilitate their use in the assessment of patients with acquired brain lesions (Rorden & Karnath, 2010), conferring a significant clinical advantage to these instruments. The complexity of cancellation tasks ranges from simple line bisection (e.g., the Albert Test; Plummer, Morris, & Dunai, 2003; Vanier et al., 1990), to large arrays of targets and non-targets (e.g., Star Cancellation Test; Linden, Samuelsson, Skoog, & Blomstrand, 2005; Manly et al., 2009; Woods & Mark, 2007), targets and related distractors (e.g., Apples Test; Bickerton et al., 2011), or even symbols and letters, such as the Symbol Cancellation Test, Letter Cancellation Test and the D2 (Bates & Lemay, 2004; Jehkonen et al., 2000; Solfrizzi et al., 2002; Uttl & Pilkenton-Taylor, 2001).

One of the most widely used instruments for the diagnosis of visuospatial neglect is the *Bells Test*, developed by Gauthier, Dehaut, and Joanette (1989), and originally published in French as the *Test des Cloches*. The task consists of an array of bells and unrelated distractors, and allows for the qualitative and quantitative assessments of visuospatial neglect. In its adaptation to Brazilian Portuguese (Fonseca et al., in press), an additional version of the Bells test was developed, containing visually related distractors in addition to the unrelated ones. The instrument was developed as a more sensitive method of detecting attentional, perceptual, praxic or executive alterations in patients with less severe neurological disorders or psychiatric conditions.

To ensure the diagnostic and prognostic accuracy of neurocognitive assessment instruments, their development must be based on strict theoretical principles, and their content and psychometric properties must

be adequate for the target population. Validity and reliability are especially important in the process of test development, and help determine whether assessment instruments are adequate for their intended purpose (Bessa, 2007; Pasquali, 2007). In addition to having strong psychometric properties, assessment instruments must also be normed for different populations and conditions. This decreases false-positive rates, and is especially important when the instrument is used to distinguish between healthy populations and clinical samples (Ostrosky-Solis et al., 2007).

The psychometric properties of adapted instruments are generally investigated based on the comparison with other assessment methods. The validity of an assessment tool refers to its ability to predict behaviors representing a specific cognitive function (Bornstein, 2011; Gorin, 2007). The construct validity of cancellation tasks is often investigated through concurrent (Azouvi et al., 2006; Bickerton et al., 2011; Lee et al., 2004) or converging validation (Bates & Lemay, 2004; Uttl & Pilkenton-Taylor, 2001; Woods & Mark, 2007). Evidence of these forms of validity is often obtained by evaluating the correlation between scores on different measures of same construct (Pasquali, 2007). Conversely, the lack of correlations between scores on a particular measure and on tasks which evaluate different constructs provide evidence of discriminant validity (Bates & Lemay, 2004; Solfrizzi et al., 2002; Uttl & Pilkenton-Taylor, 2001). The reliability of cancellation tasks is usually assessed through test-retest methods (Bickerton et al., 2011; Lee et al., 2004) or inter-rater agreement (Fabrigoule, Lechevallier, Crasborn, Dartigues, & Orgogozo, 2003; Manly et al., 2009).

Although the Bells Test is widely used in the assessment of visuospatial neglect, its validity and reliability have not been directly investigated in the literature (Menon & Korner-Bitensky, 2004). Preliminary norms are available for the Bells Test (Gauthier et al., 1989), and its sensitivity and specificity have been evaluated in the literature (Oliveira, Calvette, Pagliarin, & Fonseca, 2016; Vanier et al., 1990). The construct validity of the test has also been indirectly assessed in studies such as that of Azouvi et al. (2002), who used it as a gold-standard against which to validate other assessment instruments (Linden et al., 2005), and studies in which performance on the Bells Test was compared to that observed in other test batteries (Azouvi et al., 2003, 2006). The test has also been used as a reference for the interpretation of other cancellation tests (Rorden & Karnath, 2010; Suchan, Rorden, & Karnath, 2012). Instruments such as the Apples Test, the Character-Line Bisection Task (CLBT), the Computerized Visual Search Test and the D2 have been previously used to evaluate the validity of other cancellation tasks (Bates & Lemay, 2004; Bickerton et al., 2011; Lee et al., 2004), and instruments such as the CLBT, the Letter Cancellation Test and Star Cancellation Test have been used to assess the reliability

of cancellation tests (Lee et al., 2004; Manly et al., 2009; Uttl & Pilkenton-Taylor, 2001).

Given the relevance of the Bells Test for the assessment of attention and related processes, the assessment of its validity and reliability is an important undertaking, which may contribute to both clinical practice and research in neuropsychology. The development of an additional version of the Bells Test will also contribute to the assessment of attentional deficits in several clinical and psychiatric populations. As such, the present study had a two-fold objective: 1. To examine the convergent validity of the Bells Test by comparing it to other tests which evaluate similar constructs, and 2. To investigate the reliability of the Bells Test using the test-retest method. Our results may help elucidate the underlying mechanisms of performance on the Bells Test, and their relationship with other cognitive functions. We hypothesized that performance on the Bells Test would be positively correlated with scores on other measures of attention and visual search. Additionally, we expected that the test would show some stability in performance over time, as evidenced by positive correlations between test and retest scores.

Method

Sample

The sample was composed of 66 healthy adults ($n=44$ women) aged between 19 and 75 years, with at

least five years of formal education. All participants were native Brazilian Portuguese speakers. The sample was recruited from the states of São Paulo (SP, $n=37$) and Rio Grande do Sul (RS, $n=29$), Brazil. Participants from São Paulo had a mean age of 39.45 ± 13.15 years, and an average of 13.79 ± 3.06 years of education. Individuals recruited from Rio Grande do Sul had a mean age of 42.68 ± 15.92 years, and 14.35 ± 5.81 years of education. Since these values did not significantly differ, participants were pooled into a single sample. The sample was recruited by convenience from university, work and community settings. All participation was voluntary and preceded by written informed consent (Research Ethics Committee, protocol number 09/04908). The following inclusion criteria were applied: absence of uncorrected sensory deficits, absence of psychiatric and neurological conditions, no current or prior history of substance use problems, and no signs of dementia according to the Mini Mental State Examination (MMSE; Chaves & Izquierdo, 1992). Participants with significant symptoms of depression as indicated by the Beck Depression Inventory (BDI; Cunha, 2001), or standardized scores <7 on the Vocabulary and Block Design subtests of the Wechsler Adult Intelligence Scale (WAIS-III; Nascimento, 2004) were excluded from the sample. The demographic characteristics of the sample are summarized in Table 1.

Table 1
Demographic Characteristics of the Sample

Variables	M	SD
Age (years)	41.26	17.79
Education (years)	14.11	4.77
Frequency of reading and writing habits	16.80	4.97
Socioeconomic status	27.97	6.52
MMSE	29.11	1.33
BDI	5.76	4.83
WAIS-III Vocabulary (standardized score)	11.93	2.20
WAIS-III Block Design (standardized score)	12.80	2.99

Note. M=mean, SD=standard deviation, MMSE=Mini-Mental State Examination, BDI=Beck Depression Inventory, WAIS-III=Wechsler Adult Intelligence Scales. The frequency of reading and writing habits was evaluated by self-report through a sociocultural and health questionnaire (Fonseca et al., 2012)

Materials and Procedures

Assessment instruments were administered during a single session lasting approximately one hour. The retest was administered 5.97 ± 3.91 months after the initial assessment, on average, in a session lasting approximately half an hour. A slightly smaller sample participated in the validity study, since only 49 patients were able to return for a third assessment session.

All assessment instruments were administered by trained examiners.

The main instrument used in the present study was the Bells Test (Gauthier et al., 1989). During the adaptation of the instrument, an additional version of the test was developed. The original and new version of the test will hereafter be referred to as the BT1 and BT2, respectively (Fonseca et al., in press). The BT1 consists of an

A4 page presented in landscape orientation, containing several targets (bells) and nontargets. The adapted version of the task differed slightly from the original, in that participants were asked to cancel targets rather than circle them (Gauthier et al., 1989), to reduce variation in the responses (for a review, see da Silva, Cardoso, & Fonseca, 2011). The BT2 was developed to detect more subtle attention deficits, while the BT1 is more adequate for the assessment of visuospatial neglect. Since the test was originally developed for patients with neurological conditions (Gauthier et al., 1989), the BT1 is less complex than the BT2, which is similar to the original version save for the presence of visually-related distractors (15 bells without clappers).

Both versions of the test yield the following scores: number of targets canceled, number of omissions, number of errors (canceled distractors). In the BT2, the number of visually-related distractors canceled is also noted. The tasks are also divided into two sections. In section one (T1), the participant is asked to cancel all targets in the array and return the test stimulus to the examiner when he is finished. At this point, the examiner asks the participant whether he is sure that he has canceled all bells, and returns the test to the participant, allowing him to look for and cancel any remaining targets, and once again notify the examiner when he is finished (T2). This procedure is followed for all participants, even when all targets are canceled in T1. The time taken to complete each section of the task, as well as the sum of T1 and T2, is also recorded. The test stimulus is divided into seven vertical columns, which are numbered from left to right. The column in which the first target is canceled is recorded, as is the nature of the visual search strategy used by the participant. Strategies are classified as organized (horizontal: left-to-right or right-to-left, or vertical: bottom-up or top-down, or a mix between these two strategies), or disorganized (when no logical search strategy can be identified).

Reliability

The reliability of the BT1 and BT2 was evaluated using the test-retest method. Participants were contacted for a second assessment session, and readministered both tests. A self-report questionnaire (Fonseca et al., 2012) was used to identify whether any changes in inclusion or exclusion criteria occurred between assessments. The two versions of the BT were administered as previously described.

Construct Validity

The validity of both instruments was evaluated by comparing scores on these measures to performance on other instruments which evaluated similar constructs and had been validated for use in the Brazilian population:

Concentrated Attention Test (AC-15) (Boccallandro, 2003). The test stimulus consists of three

pages, each containing 120 pairs of words and numbers. The participant is asked to identify whether the two items in each pair are identical, and is given five minutes to complete each page. The number of correct answers obtained on each page is recorded, and the scores on the three pages are compared to determine whether participants display a decrement in attentional performance over time. The internal consistency of the AC-15, as demonstrated by the correlation between its three sections, has yielded satisfactory results, with Pearson's coefficients calculated at 0.82, 0.79 and 0.91 (Alchieri, Noronha, & Primi, 2003). Since this task relies more heavily on perceptive discrimination than on focused attention per se, this instrument was used to assess discriminant validity.

Sustained Attention Test (AS) (Sisto, Noronha, Lamounier, Bartholomeu, & Rueda, 2006). This task evaluates concentration and processing speed in addition to sustained attention. The test stimulus consists of a page with a series of targets and distractors distributed along 24 rows. The participant is given 15 seconds to cancel the targets in each row. The number of correct answers, errors and omissions on the task are recorded. The accuracy achieved in the first and last three rows of the test is compared to verify whether attentional capabilities improved, worsened or remained stable over the course of the task. Psychometric studies have found the AS to be significantly correlated with the Concentrated Attention test (*Atenção Concentrada; AC*) ($r=0.51$; $p<0.001$) (Sisto et al., 2006). The instrument has also yielded reliability coefficients ranging from 0.74 to 0.95 (Sisto et al., 2006). This instrument was used as a measure of concurrent validity for the BT.

Wechsler Adult Intelligence Scales (WAIS-III) (Nascimento, 2004) – Symbol Search and Digit-symbol Coding Tests. These tasks evaluate concentration, attentional switching and psychomotor speed. Both activities rely quite heavily on motor processes. In the symbol search test, the participant is asked to identify whether two target symbols on their left are present in a row of five geometric symbols to the right. The task has a time limit of 2 minutes. Its total score is obtained by subtracting the number of incorrect responses from the total number of correct responses. The symbol search task has demonstrated satisfactory temporal stability, as evidenced by a Pearson correlation of $r=0.89$ (Nascimento, 2004). In the digit-symbol subtest, the participant is provided with a code of matched digits and symbols, and is asked to fill in the correct symbol for a series of presented digits. The temporal stability of this subtest has been estimated at $r=0.85$ (Nascimento, 2004). The score on the digit-symbol subtest corresponds to the number of correctly filled symbols in a 2-minute time period. Both the symbol search and digit-symbol coding subtests from the WAIS-III were used as measures of discriminant validity.

Data Analysis

Test and retest scores on the BT1 and BT2 were analyzed using Student's T-test for paired samples as well as paired-samples correlations. According to the Shapiro-Wilk test, scores on the BT1 and BT2 did not meet criteria for normality. As such, associations between these and other variables were examined using Spearman correlation coefficients. The parallel-forms reliability of the BT1 and BT2 was examined by verifying the correlation between scores on both versions of the test. The construct validity of the tests was investigated by assessing

Spearman correlations between scores on the BT1 and BT2, and the AS, AC-15 and Digit-Symbol Coding and Symbol Search subtests of the WAIS-III.

Results

The descriptive statistics of scores on all measures involved in the present study, as well as the possible range of scores for each variable, are summarized in Table 2.

Paired samples T-tests for test and retest scores on the BT1 and BT2 are shown in Table 3.

Table 2
Mean, Standard Deviation and Range of Scores on the BT1, BT2, AS, AC-15, and the Digit-Symbol Coding and Symbol Search subtests of the WAIS-III

Score (maximum possible score)	M	SD	Range
BT1			
Omissions (/35)	2.27	2.71	0-19
Errors (distractors) (/280)	0.00	0.00	0-0
Time	91.57	37.85	30.52-351.27
BT2			
Omissions (/35)	2.93	3.03	0-17
Errors (related distractors) (/15)	0.08	0.84	0-15
Errors (distractors) (/265)	0.00	0.00	0-0
Time	90.19	37.77	11.79-305.81
AS			
Omissions (/150)	16.84	13.27	0-60
Errors (/150)	0.94	1.17	0-4
AC-15			
Omissions at 5 minutes (/120)	43.54	25.21	0-85
Omissions at 10 minutes (/120)	36.81	25.61	0-77
Omissions at 15 minutes (/120)	37.02	27.05	0-87
Errors at 5 minutes (/120)	3.75	5.31	0-34
Errors at 10 minutes (/120)	3.88	12.32	0-89
Errors at 15 minutes (/120)	3.25	2.98	0-11
WAIS III - Symbol Search			
Raw score (/60)	28.94	9.77	10-51
WAIS III - Digit-Symbol Coding			
Raw score (/133)	58.04	24.43	13-133

Table 3
Paired Comparisons between Test and Retest Scores on the BT1 and BT2

Scores	Test M (SD)	Retest M (DS)	d	p
BT1				
Omissions	1.12 (1.61)	0.72 (1.17)	-0.283	0.070
Errors (distractors)	0.00 (0.00)	0.00 (0.00)	-	-
Time	101.23 (45.26)	83.32 (27.73)	-0.472	0.004
BT2				
Omissions	1.73 (2.00)	1.61 (2.36)	-0.062	0.632
Errors (related distractors)	0.06 (0.49)	0.03 (0.17)	-0.083	0.641
Errors (distractors)	0.00 (0.00)	0.00 (0.00)	-	---
Time	95.29 (46.47)	76.23 (27.03)	-0.49	0.001

As can be seen in Table 3, significant differences were observed in the time taken to complete the BT1 and BT2 between the first and second assessments, possibly due to learning effects. However, the number of omission and commission errors did not differ between test and retest, supporting the temporal stability of these scores. This finding is especially important given the relevance of omission scores in the assessment of hemineglect and attentional impairments.

The analysis of paired-samples correlations between test and retest scores revealed significant moderate correlations between omission errors ($r = .454, p < .001$) and the time taken to complete the BT2 ($r = .316, p = .010$). There was also a nearly-significant association between the number of omission errors on the two administrations of the BT1 ($r = .235, p = .059$).

The paired-samples correlations between the number of omission errors and time taken to complete the BT1 and BT2 were also analyzed as a test of construct/concurrent validity. Both findings were highly significant, with correlations measured at $r = .470$ ($p < .001$) for the number of omissions on the BT1 and BT2, and $r = .732$ ($p < .001$) for the time taken to complete both tests. These values are indicative of moderate and strong associations between these variables, respectively.

The concurrent or convergent validity of the BT1 and BT2 was then examined using Spearman correlations between these measures and additional tests of attention. The results of these analyses are summarized in Table 4.

Table 4
Spearman Correlations between Performance on the BT1, BT2, AS, AC-15, and the Digit-Symbol Coding and Symbol Search subtests of the WAIS-III

Scores	BT1		BT2	
	Omissions	Time	Omissions	Time
AS				
Omissions	-0.013	0.243*	0.337*	0.283*
AC-15				
Errors at 5 min	0.408**		0.144	
Errors at 10 min	0.254*		0.182	
Errors at 15 min	0.454**		0.316*	
Correct answers at 5 min		-0.285*		-0.351**
Correct answers at 10 min		-0.247		-0.335**
Correct answers at 15 min		-0.282*		-0.323*
WAIS III - Symbol Search				
Raw Score	0.051	-0.275*	0.025	-0.280*
WAIS III - Digit-Symbol Coding				
Raw Score	0.086	-0.341*	0.175	-0.277*

Note. $p = *0.05, **0.001$

The number of omission errors on the BT1 was significantly associated with errors in the AC-15 at all time points. A similar result was obtained, but only for a single time point, with the BT2. The time taken to complete both the BT1 and BT2 was significantly correlated with all measures of attention, providing important evidence of the convergent validity of these measures.

Discussion

The present study sought to examine the convergent validity of the BT by comparing it to other measures of selective attention and visual search, and investigate its test-retest reliability in the Brazilian population. Both versions of the BT were psychometrically robust, and

were accurate in measuring concentrated and selective attention, motor skills and executive functions.

Test-retest reliability is the most commonly used method to assess the reliability of cancellation tests (Bickerton et al., 2011; Hartman-Maeir, Harel, & Katz, 2009; Wong, Cotrena, Cardoso, & Fonseca, 2010). In the present study, only the time taken to complete the BT1 and BT2 differed significantly between the two administrations of the test. The number of omission errors – the most traditionally used outcome measure for these instruments – remained stable over time. This suggests that learning effects may influence the speed with which participants complete these instruments, but not their accuracy, so that the BT1 and BT2 may be used to monitor changes in attentional performance over time. Moderate to strong correlations were found

between the number of omissions and time taken to complete the tests on both occasions, and no associations were found between the number of errors made by patients on the first and second administrations of the task. Significant associations were also found between all scores on the two versions of the BT save for the number of errors, which was very low on both tests. The correlation between scores on the BT1 and BT2 produced strong evidence of the reliability of both tests. Parallel-forms reliability is also a widely accepted method for the psychometric evaluation of cancellation instruments given the similarity in their scoring systems (Uttl & Pilkenton-Taylor, 2001). The presence of moderate correlations between test and retest scores is also an important indicator of reliability, as is the fact that no significant differences were found between test and retest scores on the BT1. The lack of previous studies of the validity of the BT1 limits the comparison of the present study with the literature (Menon & Korner-Bitensky, 2004).

The construct/convergent validity of the BT1 and BT2 were evaluated based on the correlation between scores on these measures and on the AS, AC-15 and WAIS-III Symbol search and Digit-symbol coding subtests. These instruments have all been validated for use in the Brazilian population. The correlation between multiple measures of the same construct is one of the most widely used technique to assess the validity of different assessment instruments (Azouvi et al., 2006; Lee et al., 2004; Woods & Mark, 2007). According to a recent systematic review, the comparison to gold-standard criteria is the most widely used method of evaluating the construct validity of cancellation instruments (Cotrena & Alegre, 2012).

The moderate positive correlations observed between the number of omissions in the AS, AC-15 and BT1 attest to the concurrent validity of the latter. Surprisingly, weak to nonexistent correlations were found between accuracy in the BT2 and in other measures of attention (such as the WAIS-III subtests and the accuracy in the AC-15). However, similar phenomena have been previously reported in the literature, and have been attributed to discrepancies between the attentional processes underlying performance in different tests. In light of such findings, some authors have highlighted the importance of a careful evaluation of the types of attention involved in different assessment measures (Castro, Rueda & Sisto, 2010). In the present study, it is also possible that the absence of correlations between errors on the BT and on other measures of attention was caused by the low frequency of errors on the BT1 and BT2 in the present sample. Commission errors are quite rare in both the BT1 and the BT2, in which participants are far more likely to cancel visually related distractors than unrelated ones. The absence of correlations between the number of errors on different

cancellation tasks due to ceiling effects have also been reported in the literature (Bates & Lemay, 2004).

Weak negative correlations were also found between the time required to complete the BT1 and BT2, and accuracy in the WAIS-III Digit-symbol coding. This was an unexpected finding, since the latter instrument was originally intended to provide a measure of discriminant validity. The presence of correlations between scores on these tasks suggested that Digit-symbol coding also requires a significant degree of focused attention. These findings may be interpreted as evidence of the convergent validity of the BT, whose underlying constructs also appear to be involved in the execution of tasks with a heavier reliance on fine motor skills.

The weak association between accuracy on the BT1 and BT2 and performance on WAIS-III subtests confirms the discriminant validity of the tasks. The significant correlations between the number of omissions in the BT1 and BT2 and the AS and AC-15 speak to the concurrent validity of the tests. Unfortunately, none of the tests used in the present study can be considered gold-standard techniques for the assessment of visuospatial neglect, since they were not developed for neurological populations and have not been sufficiently explored in these samples. Instruments such as the Star Cancellation Test (Linden et al., 2005; Manly et al., 2009) or the Apples Test (Bickerton et al., 2011) would have been more adequate gold-standards for the present study. However, they have not been adapted for use in Brazilian populations.

The present results support the reliability and construct validity of the BT1 and BT2, confirming their applicability to Brazilian populations. Although several other instruments allow for the assessment of attention in neurological populations, the BT is unique in its ability to also evaluate processing speed and executive functions. The location of the first target canceled also allows for a more precise assessment of visuospatial neglect. This test has not been sufficiently evaluated in healthy adults (Woods & Mark, 2007), and normative values for the BT1 or BT2 have therefore not been developed. There is, as such, a pressing need for further studies of the use of these instruments in both healthy and clinical populations to evaluate their potential in the detection of attentional impairments associated with different conditions. Qualitative and quantitative analysis of performance on these tests may also contribute to their precision and accuracy. One limitation of the present study was the fact that test-retest reliability and concurrent/discriminant validity were not evaluated in clinical samples. Such a procedure may have led to stronger correlations between test and retest scores. Future studies should seek to explore the psychometric properties of these tests in samples of different ages, education levels or with different clinical conditions (including visuospatial neglect), to confirm the statistical validity of the BT1 and BT2.

References

- Alchieri, J. C., Noronha, A. P. P., & Primi, R. (2003). Atenção Concentrada – AC15. In *Guia de referência: Testes psicológicos comercializados no Brasil* (1st ed., pp. 29-30). São Paulo: Casa do Psicólogo.
- Alqahtani, M. M. J. (2015). Assessment of spatial neglect among stroke survivors: A neuropsychological study. *Neuropsychiatry I Neuropsicologia*, 10(3-4), 95-101.
- 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 Neurology and Neuroscience*, 24(4-6), 273-285.
- Azouvi, P., Olivier, S., De Montety, G., Samuel, C., Louis-Dreyfus, A., & Tesio, L. (2003). Behavioral assessment of unilateral neglect: Study of the psychometric properties of the Catherine Bergego Scale. *Archives of Physical Medicine and Rehabilitation*, 84(1), 51-57. doi: 10.1053/apmr.2003.50062
- Azouvi, P., Samuel, C., Louis-Dreyfus, A., Bernati, T., Bartolomeo, P., Beis, J. M., ... Rousseaux, M. (2002). Sensitivity of clinical and behavioural tests of spatial neglect after right hemisphere stroke. *Journal of Neurology, Neurosurgery, and Psychiatry*, 73(2), 160-6. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1737990&tool=pmcentrez&rendertype=abstract>
- Bates, M. E., & Lemay, E. P. (2004). The d2 Test of attention: Construct validity and extensions in scoring techniques. *Journal of the International Neuropsychological Society: JINS*, 10(3), 392-400. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15147597>
- Bessa, N. M. (2007). Validade: o conceito, a pesquisa, os problemas de provas geradas pelo computador. *Estudos em Avaliação Educacional*, 18(37), 115-156. doi: 10.18222/eac183720072093
- 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. doi: 10.1037/a0023501
- Boccallandro, E. R. (2003). *Teste de Atenção Concentrada AC-15*. São Paulo: Vetor.
- Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, 23(2), 532-44. doi: 10.1037/a0022402
- Buxbaum, L. J., Dawson, A. M., & Linsley, D. (2012). Reliability and validity of the Virtual Reality Lateralized Attention Test in assessing hemispatial neglect in right-hemisphere stroke. *Neuropsychology*, 26(4), 430-41. doi: 10.1037/a0028674
- Castro, N. R. de, Rueda, F. J. M., & Sisto, F. F. (2010). Evidências de validade para o Teste de Atenção Alternada - TEALT. *Psicologia em Pesquisa*, 4(1), 40-49.
- Chaves, M. L. F., & Izquierdo, I. (1992). Differential diagnosis between dementia and depression: A study of efficiency increment. *Acta Neurologica Scandinavica*, 85(6), 378-382. doi: 10.1111/j.1600-0404.1992.tb06032.x
- Cotrena, C., & Alegre, P. (2012). Revisão Evidências de validade e fidedignidade em instrumentos de cancelamento. *Ciências & Cognição*, 17(2), 155-167.
- Cunha, J. A. (2001). *Manual da versão em português das escalas Beck*. São Paulo: Casa do Psicólogo.
- Fabrigoule, C., Lechevallier, N., Crasborn, L., Dartigues, J. F., & Orgogozo, J. M. (2003). Inter-rater reliability of scales and tests used to measure mild cognitive impairment by general practitioners and psychologists. *Current Medical Research and Opinion*, 19(7), 603-608. doi: 10.1185/030079903125002298
- Fonseca, R. P., Cardoso, C. O., Zazo, K. O., Parente, M. A. M. P., Joannette, Y., & Gauthier, L. (2018). *Teste de Cancelamento dos Sinos – TCS-1 / TCS-2*. Campinas: Vetor Editora.
- Fonseca, R. P., Zimmermann, N., Pawlowski, J., Oliveira, C. R., Gindri, G., Scherer, L. C., ... Parente, M. A. de M. P. (2012). Métodos em avaliação neuropsicológica: pressupostos gerais, neurocognitivos, neuropsicolinguísticos e psicométricos no uso e desenvolvimento de instrumentos. In J. Landeira-Fernandez & S. S. Fukusima (Eds.), *Métodos de pesquisa em neurociência clínica e experimental*. (pp. 266-296). São Paulo: Manole.
- Gauthier, L., Dehaut, F., & Joannette, Y. (1989). The Bells Test: A quantitative and qualitative test for visual neglect. *International Journal of Clinical Neuropsychology*, 11(2), 49-54.
- Gorin, J. S. (2007). Reconsidering Issues in Validity Theory. *Educational Researcher*, 36(8), 456-462. doi: 10.3102/0013189X0731
- Hartman-Maeir, A., Harel, H., & Katz, N. (2009). Kettle Test - A brief measure of cognitive functional performance: Reliability and validity in stroke rehabilitation. *American Journal of Occupational Therapy*, 64(5), 592-599. doi: 10.5014/ajot.63.5.592
- Jehkonen, M., Ahonen, J. P. P., Dastidar, P., Koivisto, A. M. M., Laippala, P., Vilkkii, J., ... Molnar, G. (2000). Visual neglect as a predictor of functional outcome one year after stroke. *Acta Neurologica Scandinavica*, 101(3), 195-201. doi: 10.1034/j.1600-0404.2000.101003195.x
- Lee, B. H., Kang, S. J., Park, J. M., Son, Y., Lee, K. H., Adair, J. C., ... Na, D. L. (2004). The Character-line Bisection Task: A new test for hemispatial neglect. *Neuropsychologia*, 42(12), 1715-1724. doi: 10.1016/j.neuropsychologia.2004.02.015
- Linden, T., Samuelsson, H., Skoog, I., & Blomstrand, C. (2005). Visual neglect and cognitive impairment in elderly patients late after stroke. *Acta Neurologica Scandinavica*, 111(3), 163-8. doi: 10.1111/j.1600-0404.2005.00391.x
- 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, benefits, and normative data. *Neuropsychology*, 23(4), 519-528. doi: 10.1037/a0015413
- Menon, A., & Korner-Bitensky, N. (2004). Evaluating unilateral spatial neglect post stroke: working your way through the maze of assessment choices. *Topics in Stroke Rehabilitation*, 11(3), 41-66. doi: 10.1310/KQWL-3HQ-4KNNM-5F4U
- Nascimento, E. (2004). *WAIS-III: Escala de inteligência Wechsler para adultos*. São Paulo: Casa do Psicólogo.
- Oliveira, C. R., Calvette, L. de F., Pagliarin, K. C., & Fonseca, R. P. (2016). Use of bells test in the evaluation of the hemineglect post unilateral stroke. *Journal of Neurology and Neuroscience*, 7.
- Ostrosky-Solis, F., Esther Gomez-Perez, M., Matute, E., Rosselli, M., Ardila, A., & Pineda, D. (2007). Neuropsi attention and memory: A neuropsychological test battery in Spanish with norms by age and educational level. *Applied Neuropsychology*, 14(3), 156-170. doi: 10.1080/09084280701508655
- Pasquali, L. (2007). Validade dos testes psicológicos: será Possível reencontrar o caminho? *Psicologia: Teoria e Pesquisa*, 23(n/s), 99-107.
- Plummer, P., Morris, M. E., & Dunai, J. (2003). Assessment of unilateral neglect. *Physical Therapy*, 83(8), 732-740. doi: 10.1093/ptj/83.8.732
- Rorden, C., & Karnath, H.-O. O. (2010). A simple measure of neglect severity. *Neuropsychologia*, 48(9), 2758-2763. <https://doi.org/10.1016/j.neuropsychologia.2010.04.018>

- Sisto, F. F., Noronha, A. P. P., Lamounier, R., Bartholomeu, D., & Rueda, F. J. M. (2006). *Testes de Atenção Dividida e Sustentada*. São Paulo: Vetor Editora.
- Silva, R. F. C., Cardoso, C. de O., Fonseca, R. P., & Alegre, P. (2011). A escolaridade no processamento atencional examinado por testes de cancelamento : uma revisão sistemática, *Ciências & Cognição*, 16(1), 180-192.
- Solfrizzi, V., Panza, F., Torres, F., Capurso, C., D'Introno, A., Colacicco, A. M., & Capurso, A. (2002). Selective Attention Skills in Differentiating between Alzheimer's Disease and Normal Aging. *Journal of Geriatric Psychiatry and Neurology*, 15(2), 99-109. doi: 10.1177/089198870201500209
- Suchan, J., Rorden, C., & Karnath, H.-O. (2012). Neglect severity after left and right brain damage. *Neuropsychologia*, 50(6), 1136-41. doi: 10.1016/j.neuropsychologia.2011.12.018
- Toglia, J., & Cermak, S. A. (2007). Dynamic assessment and prediction of learning potential in clients with unilateral neglect. *The American Journal of Occupational Therapy: Official Publication of the American Occupational Therapy Association*, 63(5), 569-579.
- Uttl, B., & Pilkenton-Taylor, C. (2001). Letter cancellation performance across the adult life span. *The Clinical Neuropsychologist*, 15(4), 521-530. doi: 10.1076/clin.15.4.521.1881
- Vanier, M., Gauthier, L., Lambert, J., Pepin, E. P., Robillard, a., Dubouloz, C. J., ... Joannette, Y. (1990). Evaluation of left visuospatial neglect: Norms and discrimination power of two tests. *Neuropsychology*, 4(2), 87-96. doi: 10.1037/0894-4105.4.2.87
- Wong, C. E. I., Cotrena, C., Cardoso, C., & Fonseca, R. P. (2010). Memoria visual: Relación con factores sociodemográficos. *Revista Mexicana de Neuropsicología*, 5(1), 10-18.
- Woods, A. J. A., & Mark, V. W. V. (2007). Convergent validity of executive organization measures on cancellation. *Journal of Clinical and Experimental Neuropsychology*, 29(7), 719-23. doi: 10.1080/13825580600954264

recebido em novembro de 2016
aprovado em maio de 2017

Sobre os autores

Cristina Elizabeth Izábal Wong é psicóloga, mestre e doutora em Psicologia, além de professora da Faculdade de Psicologia da Universidade Autónoma de Sinaloa (UAS).

Laura Damiani Branco é psicóloga pela University of Calgary, , mestre e doutoranda pelo Programa de Pós-Graduação em Psicologia PUCRS.

Charles Cotrena é psicólogo, mestre em Psicologia, especialista em Terapia Cognitivo-Comportamental e doutorando em Psicologia pela Pontifícia Universidade Católica do Rio Grande do Sul.

Yves Joannette é doutor em Neurociências e professor da Universidade de Montreal.

Rochelle Paz Fonseca é doutora em Psicologia, professora titular da Faculdade de Psicologia e do Programa em Pós-Graduação em Psicologia da PUCRS e coordenadora do Grupo Neuropsicologia Clínica e Experimental (GNCE, PUCRS).