OXFORD
UNIVERSITY PRESS

Archives of Clinical Neuropsychology 33 (2018) 66–78

Archives
of
CLINICAL
NEUROPSYCHOLOGY

Factor Structure and Measurement Invariance of the Coimbra Neuropsychological Assessment Battery (BANC)

Octávio Moura[1,2,*], Cristina P. Albuquerque[1,3], Maria Salomé Pinho[1,3,4], Manuela Vilar[1,2,3],
Ana Filipa Lopes[1,2], Isabel Alberto[3], Marcelino Pereira[1,2,3], Maria João Seabra Santos[1,2,3],
Mário R. Simões[1,2,3]

[1]*Neurosciences, Neuropsychology and Cognitive Assessment group from the Cognitive and Behavioral Center for Research and Intervention (CINEICC), University of Coimbra, Portugal*
[2]*Psychological Assessment and Psychometrics Laboratory, Faculty of Psychology and Educational Sciences, University of Coimbra, Portugal*
[3]*Faculty of Psychology and Educational Sciences, University of Coimbra, Portugal*
[4]*Memory, Language and Executive Functions Laboratory, Faculty of Psychology and Educational Sciences, University of Coimbra, Portugal*

*Corresponding author at: Psychological Assessment and Psychometrics Laboratory, Faculty of Psychology and Educational Sciences, University of Coimbra, Rua do Colégio Novo, 3000-115 Coimbra, Portugal. Tel.: + 351-239851450; fax: +351-239851462.
E-mail address: octaviomoura@gmail.com (O. Moura).

Editorial Decision 5 May 2017; Accepted 11 May 2017

Abstract

Objective: This study aimed to investigate the factor structure of the Coimbra Neuropsychological Assessment Battery (BANC), which is an individually administered battery designed to assess a wide range of neurocognitive functions in children.
Method: Using the standardization sample of the BANC, a confirmatory factor analysis and a multiple-group analysis were conducted to examine the factor structure and the measurement invariance of three main domains (Memory, Language, and Attention/Executive Functions) in 833 children aged 7–15 years.
Results: Consistent with the BANC's conceptualization, the three-correlated-factor model demonstrated the most adequate fit to the data. The measurement invariance of the three-correlated-factor model across two age-groups (7–9 years and 10–15 years) was supported (configural, metric, and partial scalar invariance).
Conclusion: Overall, the BANC shows adequate psychometric properties and provides useful information regarding the children's neuropsychological functioning.

Keywords: Coimbra Neuropsychological Assessment Battery (BANC); Neuropsychological assessment; Confirmatory factor analysis; Measurement invariance

---

Introduction

In recent decades, specific neuropsychological batteries were developed to provide a comprehensive evaluation of the diverse neuropsychological functions in children, adolescents and adults. Although neuropsychological assessment led to major conceptual and theoretical advances in the understanding of normal and abnormal patterns of brain-behavior relations, in certain cases, their psychometric properties are objectionable or are not clearly noted. Many of the practices in neuropsychology have been criticized (see for a review: Reynolds & Mason, 2009) for a lack of attention to certain principles of research design and the failure to incorporate the many advances in psychometric methods of the last years [e.g., confirmatory factor analysis (CFA), measurement invariance analysis, exploratory structural equation modeling].

Indeed, few studies conducted exploratory factor analysis (EFA) or CFA for the neuropsychological batteries most widely used in pediatric age, which clearly contrast with the large number of factor studies developed with intelligence scales for children (e.g., Wechsler Intelligence Scale for Children – WISC; Canivez, Watkins, & Dombrowski, 2017; Donders, Elzinga,

doi:10.1093/arclin/acx052

Kuipers, Helder, & Crawford, 2013; Weiss, Keith, Zhu, & Chen, 2013). Certain exceptions include Stinnett and colleagues (2002), who conducted an EFA for Developmental Neuropsychological Assessment (NEPSY) and found that the one-factor model was the most interpretable and parsimonious factor solution. This one-factor model accounted only for 24.93% of the variance, with factor loadings ranged between .24 and .64. Using the standardization sample, Mosconi, Nelson and Hooper (2008) performed a CFA for NEPSY and found that the theoretically derived five-factor model was not statistically admissible because it produced negative error variance. The authors found that a four-factor model (without the Executive Function/ Attention domain) yielded an adequate model fit for the entire sample [Goodness-of-Fit Index (GFI) = .95, Standardized Root Mean Square Residual (SRMR) = .04 and Root Mean Square Error of Approximation (RMSEA) = .04], for both the younger (GFI = .97, SRMR = .04, and RMSEA = .04) and older children (GFI = .91, SRMR = .05, and RMSEA = .06), with factor loadings ranged between 0.31 and 0.69 for the entire sample.

Exploratory factor analytic studies have also produced different factor solutions for Halstead-Reitan Neuropsychological Battery for Older Children (see for a review: Ross, Allen, & Goldstein, 2014). For example, Krug, Dean and Anderson (1995) found a four-factor solution (Speed of Operation, Tactile-Motor Integration, Attention, and Visuo-Spatial Memory) in a sample of 800 children with learning disabilities, whereas Livingston, Gray, Haak and Jennings (1997) identified a seven-factor solution (Spatial Processing Speed; Motor Strength; Nonverbal Learning and Memory; Visual, Auditory, Somesthetic Sensation; Auditory Processing; Motor Speed; and Visual Attention) that accounted for 76% of the variance in a sample of 516 children (aged 9–14 years) with academic and behavioral concerns.

Taken together, these findings showed that the factor structure derived from EFA and/or CFA are moderately different than the proposed theoretical models. These results are not surprising because these batteries were derived theoretically and were not based on factor analysis. Although a conceptual framework is important during the development of neuropsychological test batteries, adequate psychometric properties (e.g., factor structure, reliability, measurement error, temporal stability, sensitivity, specificity, diagnostic accuracy) are essential for the valid use in clinical practice and for a more accurate interpretation of test scores (Reynolds & Mason, 2009; Strauss, Sherman, & Spreen, 2006). On the other hand, neurocognitive functions are subject to distinct developmental trajectories, which may also explain the variability in the factor structure found in the pediatric neuropsychological batteries. Neurodevelopmental studies have shown that the adult-level performance on the most complex executive functions does not occur until adolescence or even early adulthood, which is consistent with the view that frontal lobes are the latest brain structures to mature (Best & Miller, 2010; Korkman, Lahti-Nuuttila, Laasonen, Kemp, & Holdnack, 2013). Working memory capacity has been found to gradually develop throughout childhood and into young adulthood (Gathercole, Pickering, Ambridge, & Wearing, 2004; Huizinga, Dolan, & van der Molen, 2006), whereas performance on language measures (i.e., phonological processing and comprehension of instructions) seems to significantly improve until age of 9 years (Korkman et al., 2013). Korkman and colleagues (2013) using the North American standardization sample of the NEPSY-II found that neurocognitive measures development is rapid in the age range 5–9 years followed by a deceleration in the rate of development. They also observed that peak performances were reached at 14–16 years, except for measures tapping executive functions, verbal memory and visuospatial performance that continue to develop beyond the age of 16 years.

It is also important to note that differences in performance on neurocognitive measures may also be related to specific linguistic characteristic and the socioeconomic status, mainly when comparing individuals from different languages or cultures. Indeed, the level of orthographic consistency of a language can influence the performance of some verbal neurocognitive measures (e.g., phonological awareness tasks may be more difficult in opaque than in transparent orthographies); whereas word, letter or digit span tasks may be more/less dependent on working memory based on phonemic structure (Lobley, Baddeley, & Gathercole, 2005; Pickering, 2004; Vaessen et al., 2010). The socioeconomic status may also have an affect on the performance of some neurocognitive measures, specifically in culturally heterogeneous contexts where some individuals may not fit the standardization sampling characteristics (Shuttleworth-Edwards, 2016).

More recently, Wicherts (2016) highlighted the importance of conducting measurement invariance analysis, in addition to the factor analysis, when norming a neurocognitive battery. The measurement invariance analysis is relevant to analyze if the factor model is valid throughout the standardization sample or for specific subgroups (e.g., based on age, gender, socioeconomic status, ethnicity, and educational background) in which the battery is used. If the measurement invariance is not achieved, the latent neurocognitive ability that the subtest is supposed to measure cannot explain all observed group differences on that subtest, which negatively affects the quality of assessment and decisions made based on the subtest scores.

Despite the fact that Portuguese is the sixth most natively spoken language in the world with more than 200 million native speakers (Lewis, Simons, & Fennig, 2015), there are few normalized neuropsychological tests for the evaluation of children. Thus, the objective of the current study was to evaluate the factor structure of a new neuropsychological assessment battery for children, named the Coimbra Neuropsychological Assessment Battery (BANC; Simões et al., 2016). Although BANC has six theoretically derived domains, our interest is solely in the evaluation of the Memory, Language and Attention/Executive Functions domains. Laterality, Motor Function and Orientation domains were excluded because the first is only an observation task that does not yield an age-adjusted-scaled score, and the two others are outside the psychometric interest for this study. Specifically, our purpose was to test the BANC's factor structure that was theoretically derived (three-correlated-factor model) and three alternative models (three-uncorrelated-factor model, four-correlated-factor model, and five-correlated-factor model) through a CFA. Because three subtests of the BANC have different tasks for children aged 7–9 years and 10–15 years (more complex tasks for children aged 10–15 years), we additionally conducted a multiple-group analysis (measurement invariance) to evaluate whether the factor structure of BANC would be equivalent across these two age-groups. We hypothesized that: (1) the three-correlated-factor model (BANC's theoretical model) would be the model that best represents the underlying structure of the BANC; and (2) the three-correlated-factor model would operate equivalently across children aged 7–9 years and 10–15 years.

Coimbra Neuropsychological Assessment Battery

The BANC (Simões et al., 2016) is an individually administered battery designed to assess a wide range of neurocognitive functions in children ranging from 5 to 15 years old. This battery is the first Portuguese neuropsychological assessment battery that taps different functions of children's neuropsychological development; it includes 15 subtests that are organized into six theoretically derived domains: (1) Memory; (2) Language; (3) Attention/Executive Functions; (4) Motor Function; (5) Laterality; and (6) Orientation. Table 1 highlights the six domains, the 15 subtests that are grouped within each domain, the description and the different scores of each subtest. Some subtests have different tasks depending on the child's age. The administration time of the entire battery is approximately 120 min. All of the subtests' raw scores are converted into age-adjusted-scaled scores ($M = 10$; $SD = 3$). In addition to the individual age-adjusted-scaled scores for each subtest, the BANC yields three domain scores ($M = 100$; $SD = 15$; and percentile rank) for Memory, Language, and Attention/Executive Functions domains.

The results provided by the BANC may yield relevant information to assist the clinician in identifying the pattern of neuropsychological strengths and weakness in typically developing children and in children with neurodevelopmental disorders. The clinical and diagnostic utility of the BANC have been observed in studies with children with developmental dyslexia (Moura, Moreno, Pereira, & Simões, 2015; Moura, Simões, & Pereira, 2015a, 2015b), epilepsy (Lopes, Monteiro, Fonseca, Robalo, & Simões, 2014; Lopes, Simões, & Leal, 2014), specific language impairment (Coelho, Albuquerque, & Simões, 2013), oppositional defiant disorder (Sá, Albuquerque, & Simões, 2008), traumatic brain injury (Santos, 2006) and attention-deficit hyperactivity disorder (Moura et al., 2017). For example, the phonological awareness revealed a high diagnostic accuracy in the developmental dyslexia (sensitivity = 93.8%, specificity = 94.1%, AUC values from the ROC curve analysis ≥ .950), and naming speed in the attention-deficit hyperactivity disorder (sensitivity = 75%, specificity = 88.2%, AUC values from the ROC curve analysis ≥ .825) (Moura, Moreno, et al., 2015; Moura et al., 2017). These data suggest an adequate discriminant validity of the BANC.

Reliability was obtained through test–retest stability, internal consistency and interrater reliability (Simões et al., 2016). The internal consistency was calculated for some subtests based on Cronbach's alpha and split-half. The results indicate adequate internal consistency from $\alpha = .72$ (Phonological Awareness – Substitution B) to $\alpha = .91$ (Phonological Awareness – Deletion). For the subtests in which the prior two methods were inappropriate, the authors used test–retest stability coefficients (see Table 2).

Method

Participants

The standardization sample of the BANC was used in the current study. The standardization sample is a national stratified random sample that consists of 1104 Portuguese children and adolescents between the ages of 5 and 15 years, which considered the following criteria: (1) *age* with approximately 100 individuals by age group, with a mean age of 10.01 years ($SD = 3.16$ years); (2) *gender* with the same number of boys and girls by age level; (3) *school grade* from preschool to 10th grade, with approximately 100 participants per grade; (4) *residential area* arranged by urban ($N = 781$), moderately urban ($N = 186$) and rural ($N = 137$) equal to the Portuguese organization (INE/DGOTDU, 1998); and (5) *geographic region* arranged by coastal areas ($N = 928$) and interior areas ($N = 176$) similar to the Portuguese population's organization.

Children with neurological disease, neurodevelopmental disorder, learning disabilities, psychopathology, disruptive, impulse-control, and conduct disorders, sensory deficits, one or more school retentions, special educational needs or who benefited from special education services were excluded from the BANC normative sample.

Table 1. Description of BANC subtests

Domains/subtests	Description	Scores (age)
Memory		
Word Learning List	Assesses the learning ability, retention, recall and recognition of a word list. The child begins by learning a list of 15 words (the first trial represents the Immediate Recall score) during four trials (Total Learning score). A new list with 15 words is then presented and recalled once (Interference Recall score). Then, the child is requested to recall the first word list either immediately (Short-Delay Recall score) and after a 20- to 30-min delay (Long-Delay Recall score). Finally, 45 words are presented to the child to indicate whether the words belong to the first list (Recognition score).	Immediate Recall (5–15) Total Learning (5–15) Interference Recall (5–15) Short-Delay Recall (5–15) Long-Delay Recall (5–15) Recognition (5–15)
Stories Memory	Assesses retention, recall and recognition, as well planning, organizing, sequencing and language skills. It embraces 4 stories: stories A and B are administered to children aged from 5 to 9 and stories C and D to children aged from 10 to 15. The examiner reads each story, and the child retells it immediately after having heard it (Immediate Recall score) and after a delay of 20–30 min (Delayed Recall score). Finally, the child answers multiple-choice questions regarding each story (Recognition score).	Immediate Recall AB (5–9) Delayed Recall AB (5–9) Recognition AB (5–9) Immediate Recall CD (10–15) Delayed Recall CD (10–15) Recognition CD (10–15)
Memory of Faces	This subtest assesses the recognition ability of 16 unfamiliar faces. First, the faces are shown to the child and immediately after the last face is presented, the child identifies, within sets of three faces, each one of the previously viewed faces (Immediate Recall score). After 20–30 min, the child identifies the same faces from different sets of three (Delayed Recall score).	Immediate Recall (5–15) Delayed Recall (5–15)
Rey Complex Figure	This subtest assesses a variety of cognitive processes, but its primary purpose is to assess visuospatial ability and visual memory. The child must copy the Rey Complex Figure, followed by a Short-Delay Recall (3 min after) and a Long-Delay Recall (20–30 min after).	Copy (5–15) Short-Delay Recall (5–15) Long-Delay Recall (5–15)
Corsi Blocks	It is a visuospatial short-term memory test. The examiner taps with his finger on a board with nine blocks according to prearranged sequences and the child must reproduce each of those tapping patterns.	Corsi Blocks – Immediate Recall (5–15)
Language		
Phonological Awareness	In the Deletion task, the child was asked to delete a particular phoneme from familiar words. In the Substitution task, the child was asked to repeat familiar words after having replaced one or more phonemes for another phoneme(s).	Deletion (6–15) Substitution A (6–9) Substitution B (10–15) Total score (6–15)
Comprehension of Instructions	This subtest assesses receptive language, at the semantic and syntactic level, through the child's answers to 27 oral instructions. These instructions contain several concepts (e.g., expressing quantity, sequence, temporal or spatial relations), which involve an increasing conceptual complexity level and different materials.	Comprehension of Instructions (5–15)
Naming Speed	In each naming speed subtest, the child should name, as quickly as possible, 50 visual stimuli randomly displayed on a card in a 10 × 5 matrix. The stimuli of the Rapid Automatized Naming (RAN) - Colors are yellow, blue, red, black, and green circles. The stimuli of the RAN – Number are 2, 4, 6, 7 and 9. The stimuli of the Rapid Alternating Stimulus (RAS) – Colors/Shapes are the circle, rectangle, square and triangle, which present the colors yellow, red, black and green.	RAN – Colors (5–6) RAN – Numbers (7–15) RAS – Colors/Shapes (7–15)
Attention/Executive Functions		
Cancellation	This subtest assesses selective and sustained attention. The material comprises an A3 sheet with 1600 squares arranged in lines and 2 (for children aged 5–9 years) or 3 (for children aged 10–15 years) model squares (signs) placed at the top of the sheet. The child's task consists of crossing out the squares that are equal to the model squares during 10 min. The score is determined through a formula that considers the number of squares correctly crossed, omitted and incorrectly crossed.	2 Signs (5–9) 3 Signs (10–15)
Trail	The Trail – Part A assesses visuospatial sequencing and rapid visual search. The child must draw a line connecting 25 encircled numbers randomly distributed on a sheet of paper, sequentially from 1 to 25. The Trail – Part B is more complex than part A because it has greater requirements in terms of motor speed and rapid visual search, and demand mechanisms of cognitive shifting and flexibility. The child must draw a line connecting 25 circles with numbers or letters, randomly distributed on a sheet of paper, alternating between numbers and letters (1, A, 2, B, etc.).	Trail – Part A (6–15) Trail – Part B (7–15)
Verbal Fluency	This subtest requires the mobilization of verbal skills, memory and executive functions. The child must generate as many different words as possible within a time constraint of 60 seconds, according to three semantic categories (Animals, Names and Food) and three phonemic categories (letters P, M, R).	Semantic (5–15) Phonemic (7–15)
Tower	This subtest assesses the executive functions of planning, monitoring, self-regulation and problem solving. The subtest is composed of 14 models that the child must reproduce by creating a tower with three balls of different colors (red, blue and green) and three pegs (large, medium and small). The child must move the three colored balls to specific positions on the three pegs in a specific number of moves (starting with one move and gradually increasing to five moves).	Correct First Trials (5–15) Correct Models (5–15) Total of Trials (5–15)

(*continued on next page*)

Table 1. (*continued*)

Domains/subtests	Description	Scores (age)
Motor Function	This subtest assesses manual and finger dexterity. The child must put as many pins as possible in a board with 50 holes in 30 s, first with the preferred hand (2 trials), then with the non-preferred hand (2 trials) and finally with the two hands simultaneously (2 trials).	Preferred Hand (5–15) Non-Preferred Hand (5–15) Both Hands (5–15)
Laterality	This subtest assesses the recognition of the right and left notion and the laterality dominance through the execution of a set of proposed activities.	(No Scaled Scores) (5–15)
Orientation	This subtest comprises 17 questions related to personal and temporal information.	Orientation – Total (5–15)

Table 2. Descriptive statistics, item-total correlation and test–retest

Domains and subtests	Skewness	Kurtosis	$r_{\text{i-t}}$	Test–retest
Memory				
Word Learning List (Short-Delay Recall)	−0.053	−0.301	.408	.470
Word Learning List (Long-Delay Recall)	−0.034	−0.304	.409	.675
Stories Memory (Immediate Recall)	−0.008	−0.231	.439	.835
Stories Memory (Delayed Recall)	−0.062	−0.123	.453	.790
Memory of Faces (Immediate Recall)	−0.203	−0.435	.243	.740
Memory of Faces (Delayed Recall)	−0.235	−0.467	.264	.533
Rey Complex Fig. (Short-Delay Recall)	0.019	−0.254	.432	.816
Rey Complex Fig. (Long-Delay Recall)	0.023	−0.233	.452	.807
Corsi Blocks	0.010	−0.199	.163	.602
Language				
Phonological Awareness (Deletion)	−0.332	−0.452	.465	.873
Phonological Awareness (Substitution)	−0.035	−0.272	.544	.783
Comprehension of Instructions	−0.077	−0.291	.355	.744
Naming Speed (RAN – Numbers)	−0.388	0.088	.321	.802
Naming Speed (RAS – Colors/Shapes)	−0.514	0.276	.462	.863
Attention/Executive Functions				
Cancellation	−0.069	−0.099	.330	.759
Trail – Part A	−0.487	0.189	.393	.745
Trail – Part B	−0.663	0.513	.427	.528
Verbal Fluency (Semantic)	0.004	−0.226	.330	.762
Verbal Fluency (Phonemic)	0.005	−0.268	.334	.386
Tower (Total of Trials)	−0.241	−0.375	.144	.533

Note: $r_{\text{i-t}}$ = corrected item-total correlation with the subtests of their specific domain; RAN = rapid automatized naming; RAS = rapid alternating stimulus.

To have the same number of subtests to be estimated, we excluded children who were aged 5 and 6 years from this study because three subtests (Naming Speed, Trail – Part B and Phonemic Verbal Fluency) are not administered to these age groups. Thus, the subsequent statistical analyses only consider children between the ages of 7 and 15 years. Seventy-one children were eliminated from the analysis due to missing data, resulting in a total of 833 children in the final sample.

Procedure

Voluntary participation was requested of all participants of the standardization sample, and the objectives of the study were fully explained. This research was approved by the Scientific Council of the Faculty of Psychology and Educational Sciences – University of Coimbra, the National Commission of Data Protection and the Portuguese Foundation for Science and Technology. Informed consent information was gathered from parents and from school directors. All of the subtests were individually administered in one or two test sessions (if a second testing session was necessary, it was completed within one week), in a quiet school space during a regular school day. The administration of the BANC was made by psychologists trained and experienced in neuropsychological assessment. No incentives (fees or extra credit) were offered in exchange for participation.

Statistical Analyses

Descriptive statistics, correlation analysis and test–retest reliability were conducted using IBM SPSS 20. To test the factor structure of the BANC, a CFA was performed using IBM SPSS Amos 20. The models tested were estimated through covariance matrices using maximum likelihood estimation. Model fit was assessed through a number of indices: chi-square (χ^2), Comparative Fit Index (CFI), Parsimony Comparative Fit Index (PCFI), SRMR, RMSEA and Akaike Information Criterion (AIC). Chi-square is known to be extremely sensitive to sample size, meaning that with larger samples, reasonable models are likely to produce statistically significant chi-square p values (Bentler, 1990; Bryant & Yarnold, 1995; Jöreskog & Sörbom, 1989). In these cases, the analysis of other fit indices is recommended. Two absolute fit indices were used (SRMR and RMSEA), as well as an incremental fit index (CFI) and two parsimonious fit indices (PCFI and AIC). Hu and Bentler (1999) recommend a CFI of > .95, a SRMR of <.08 and an RMSEA of <.06 to determine good fit. For PCFI, values of [.6–.8] indicate a reasonable fit and >.8 a good fit (Blunch, 2008). The AIC was used to compare models, with smaller values representing a better fit. If χ^2 is sensitive to sample size, most of the fit indices are sensitive to model complexity (i.e., number of observed variables and number of factors). As suggested by Marsh, Hau and Wen (2004) these traditional cutoff values should not be used as rules of thumb. Therefore, more stringent cutoff values are recommended for simple models, and less stringent cutoff values are recommended for more complex models (Cheung & Rensvold, 2002; Marsh et al., 2004).

Using the general procedures outlined by Byrne (2004, 2010) and Vandenberg and Lance (2000), we tested measurement invariance based on the analysis of mean and covariance structures that encompassed a series of hierarchically ordered steps that began with the establishment of a baseline model for each age-group separately (aged 7–9 years and 10–15 years), followed by tests for increasingly more stringent levels of constrained equivalence across both groups: (1) for *configural* invariance, no equality constraints were imposed on the parameters across the two groups; (2) for *metric* invariance ("weak factorial invariance"), we constrained factor loadings to be equivalent across groups; and (3) for *scalar* invariance ("strong factorial invariance"), we constrained factor loadings and intercepts to be equal across groups. To determine evidence of invariance, we compared the difference values of χ^2, *df* and CFI between the configural and the other two models (i.e., the configural model provides the baseline value against which all subsequently invariance models are compared). It is commonly accepted that evidence for invariance is obtained if (1) the multi-group model exhibits an adequate fit to the data, (2) the χ^2 difference value ($\Delta\chi^2$) is not statistically significant ($p > .05$), and (3) the CFI difference value (ΔCFI) is <−.010 (Byrne, 2010; Cheung & Rensvold, 2002; Jöreskog & Sörbom, 1996). Some studies have demonstrated that alternative fit indices are often preferable over the $\Delta\chi^2$ in the context of measurement invariance and the cutoff value for Δ goodness-of-fit indices depend of the factor structure, sample size, number of groups and constraint level (Khojasteh & Lo, 2015; Meade, Johnson, & Braddy, 2008).

Results

For the descriptive statistics, correlations, CFA and measurement invariance, the age-adjusted-scaled scores of the neuropsychological subtests of the Memory, Language and Attention/Executive Functions domains were the sole object of study (as previously noted, the Laterality, Motor Function and Orientation domains were excluded from this study). In the Memory domain, our interest is only in the immediate and delayed recall scores.

Descriptive Statistics, Item-Total Correlation and Test–Retest Reliability

Skewness and kurtosis values were examined to determine the normality of the data distribution. Curran, West and Finch (1996) suggested that values approaching 2 and 7 for skewness and kurtosis, respectively, resulted in significant problems with maximum likelihood estimation. As shown in Table 3, all of the subtests showed skewness and kurtosis values < 1, which suggested adequate distribution for maximum likelihood estimation.

The item-total correlation of each subtest with their specific domain revealed, in general, moderate correlation coefficients (see Table 2). The lowest item-total correlation coefficient was found on the Tower ($r_{\text{i-t}} = .144$), whereas the highest coefficient was found on the Phonological Awareness – Substitution ($r_{\text{i-t}} = .544$).

The test–retest sample consisted of 69 typically developing children (8 and 10 years old). Test–retest intervals ranged from 18 to 35 days, with a mean of 27.88 days ($SD = 3.79$) between administrations. Test–retest coefficients were based on the raw scores, with Pearson correlation coefficients ranged from .386 (Phonemic Verbal Fluency) to .873 (Phonological Awareness – Deletion) (see Table 2). The Language domain showed the highest stability across time (the mean of the test–retest coefficient was $r = .813$), which suggested minimal practice effects.

Confirmatory Factor Analysis

In addition to the BANC's theoretically derived three-factor model, we also tested alternative factor structures to evaluate how the verbal and the visual memory measures from the Memory domain operate separately, as well as how the attention

Table 3. Standardized factor loadings for the confirmatory bifactor model

	Confirmatory bifactor model			
	General factor	Factor 1 (Memory)	Factor 2 (Language)	Factor 3 (Attention/EF)
Word Learning List (Short-Delay Recall)	.670	.511		
Word Learning List (Long-Delay Recall)	.689	.584		
Stories Memory (Immediate Recall)	.736	.580		
Stories Memory (Delayed Recall)	.749	.593		
Memory of Faces (Immediate Recall)	.119	.075		
Memory of Faces (Delayed Recall)	.143	.022		
Rey Complex Fig. (Short-Delay Recall)	.258	.029		
Rey Complex Fig. (Long-Delay Recall)	.281	.043		
Corsi Blocks	.146	.106		
Phonological Awareness (Deletion)	.231		.396	
Phonological Awareness (Substitution)	.329		.525	
Comprehension of Instructions	.341		.465	
Naming Speed (RAN – Numbers)	.062		.247	
Naming Speed (RAS – Colors/Shapes)	.265		.391	
Cancellation	.248			.469
Trail – Part A	.154			.416
Trail – Part B	.232			.480
Verbal Fluency (Semantic)	.231			.262
Verbal Fluency (Phonemic)	.212			.275
Tower (Total of Trials)	.103			.180

Note: RAN = rapid automatized naming; RAS = rapid alternating stimulus.

and the executive functions measures operate individually from the Attention/Executive Functions domain. Thus, a CFA was performed to evaluate four factor models: *Model 1*: three-uncorrelated-factor model (Memory, Language, and Attention/Executive Functions); *Model 2*: three-correlated-factor model (Memory, Language, and Attention/Executive Functions); *Model 3*: four-correlated-factor model (Verbal Memory, Visual Memory, Language, and Attention/Executive Functions); and *Model 4*: five-correlated-factor model (Verbal Memory, Visual Memory, Language, Attention and Executive Functions). The CFA was estimated through the maximum likelihood estimation that required the assumption of multivariate normality. The univariate statistics (skewness and kurtosis) performed previously for each variable (subtests were treated as continuous indicators) and the multivariate value represented by Mardia's coefficient of multivariate kurtosis performed for each factor model showed adequate values, which are suggestive of multivariate normality in the sample.

A preliminary analysis of the modification indices for each of the factor models suggested the addition of error covariances only for the subtests that include two scores. The addition of these error covariances between the different scores of the same subtest makes statistical and empirical sense (see Boomsma, 2000; Byrne, 2010 regarding when respecification of models is appropriate) because, in general, measure the same underlying neuropsychological function and are strongly correlated (e.g., $r = .77$, $p < .001$, for short- and long-delay recall scores of the Word Learning List subtest with; $r = .63$, $p < .001$, for deletion and substitution scores of the Phonological Awareness subtest). No cross-loadings or additional error covariances between subtests were suggested by the examination of the modification indices.

The goodness-of-fit indices for Model 1 indicated a poor fit between the data and the estimated model, with $\chi^2(162) = 872.249$, $p < .001$, CFI = .880, SRMR = .125, RMSEA = .073 (90% CI = .068–.077), PCFI = .751 and AIC = 968.249. The three-correlated-factor model (Model 2) showed a good model fit, with χ^2 (159) = 396.876, $p < .001$, CFI = .960, SRMR = .045, RMSEA = .042 (90% CI = .037–.048), PCFI = .803 and AIC = 498.876. As illustrated in Fig. 1, the three factors were highly correlated and revealed, in general, adequate factor loadings. Memory and Language factors had an adequate reliability ($\alpha = .74$ and $\alpha = .72$, respectively), whereas a marginal reliability was found for the Attention/Executive Functions factor ($\alpha = 0.64$). For Model 3, a factor correlation greater than 1 (i.e., Heywood case) was found between Visual Memory and Attention/Executive Functions. Similarly, for Model 4, a factor correlation greater than 1 was also found between Attention and Executive Functions. Therefore, these two factor solutions were not statistically admissible. Taken together, the results from the CFA showed that the three-correlated-factor model provided the best fit to the data.

Confirmatory bifactor model. Because the three latent factors from Model 2 (three-correlated-factor model) were highly correlated, a confirmatory bifactor model was additionally estimated in order to analyze whether BANC could be represented simultaneously by a general neurocognitive factor and specific factors. In a confirmatory bifactor model, all factors were

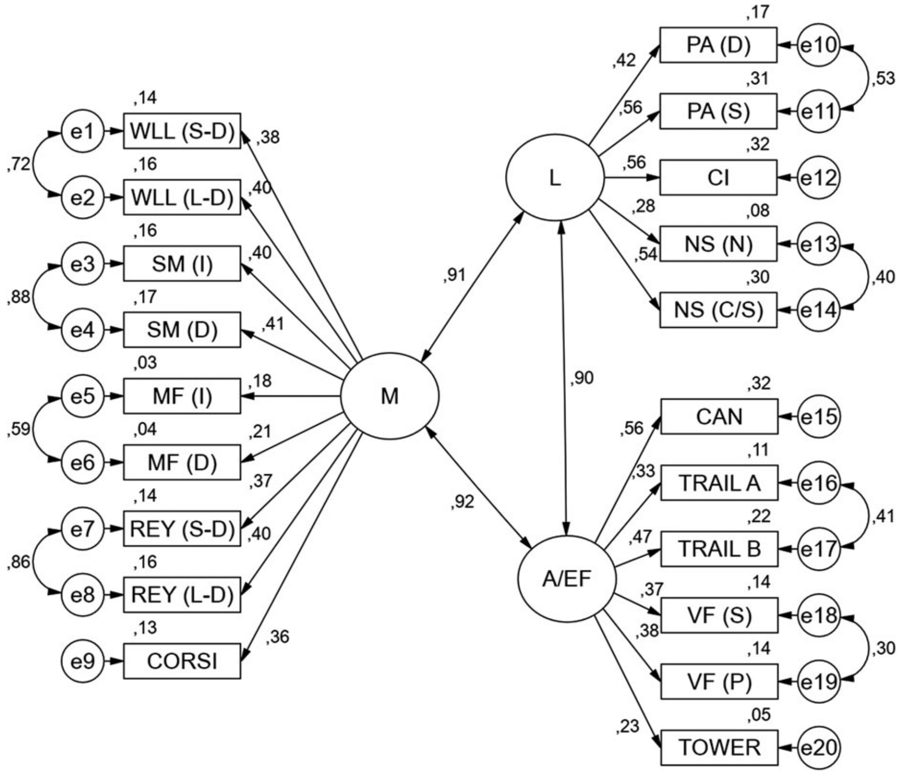

Fig. 1. BANC – three-correlated-factor model (standardized solution).
Note: Factors: M = Memory, L = Language, A/EF = Attention/Executive Functions. Observed Variables: WLL (S-D) = Word Learning List (Short-Delay Recall), WLL (L-D) = Word Learning List (Long-Delay Recall), SM (I) = Stories Memory (Immediate Recall), SM (D) = Stories Memory (Delayed Recall), MF (I) = Memory of Faces (Immediate Recall), MF (D) = Memory of Faces (Delayed Recall), Rey (S-D) = Rey Complex Figure (Short-Delay Recall), Rey (L-D) = Rey Complex Figure (Long-Delay Recall), Corsi = Corsi Blocks, PA (D) = Phonological Awareness (Deletion), PA (S) = Phonological Awareness (Substitution), CI = Comprehension of Instructions, NS (N) = Naming Speed (RAN – Numbers), NS (C/S) = Naming Speed (RAS – Colors/Shapes), CAN = Cancellation, Trail A = Trail – Part A, Trail B = Trail – Part B, VF (S) = Verbal Fluency (Semantic), VF (P) = Verbal Fluency (Phonemic), Tower (Tower – Total of Trials).

specified to be orthogonal (i.e., the correlation between the factors are restricted to zero). Thus, each indicator (neurocognitive test) is simultaneously explained by the general (neurocognitive) factor and the specific factor to which it theoretically belongs.

The confirmatory bifactor model yielded an adequate fit to the data, albeit the goodness-of-fit indices were less adequate than the Model 2, with χ^2 (144) = 610.183, $p < .001$, CFI = .921, SRMR = .077, RMSEA = .062 (90% CI = .057–.068), PCFI = .698 and AIC = 742.183. Table 3 shows the factor loadings for the general factor and specific factors.

Measurement Invariance Analysis

Because three subtests of the BANC (Stories Memory, Phonological Awareness – Substitution and Cancellation) have different tasks for children aged 7–9 years and 10–15 years (more complex tasks for children aged 10–15 years), a multiple-group analysis was conducted for the three-correlated-factor model (Model 2) to evaluate whether the factor structure of BANC would be equivalent across these two age-groups.

The three-correlated-factor model (baseline model) yielded a good fit for children aged 7–9 years: CFI = .955, SRMR = .058, RMSEA = .044 (90% CI = .033–.054), PCFI = .799, AIC = 351.928; and for children aged 10–15 years CFI = .962, SRMR = .048, RMSEA = .043 (90% CI = .036–.050), PCFI = .805, AIC = 417.260. After establishing the baseline model for each group, we tested for configural invariance in which no equality constraints were imposed on the parameters across the two groups. The configural model had adequate model fit, which suggested that both the number and pattern of factors were equivalent across groups (see Table 4). The evaluation of metric invariance was conducted by constraining the factor loadings (regression slopes) to be equivalent across groups. The $\Delta\chi^2$ (12) = 13.778, $p = .315$ and ΔCFI = −.001 values indicated that the invariance of factor loadings did not result in a significantly worse model fit compared with configural invariance, which supported metric invariance. Scalar invariance was examined by constraining factor loadings and intercepts to be

Table 4. Measurement invariance analysis

	CFI	SRMR	RMSEA (90% CI)	χ^2	*df*	Δdf	$\Delta\chi^2$	ΔCFI
Configural	.960	.059	.031 (.026–.035)	565.255	318			
Metric	.958	.064	.030 (.026–.034)	590.153	335	17	24.898, $p = .097$	−.002
Scalar	.941	.064	.035 (.031–.039)	719.123	355	37	153.868, $p < .001$	−.019
Scalar (partial)	.960	.064	.029 (.024–.032)	593.156	354	36	27.901, $p = .831$	.000

Note: CFI = Comparative Fit Index; SRMR = standardized root mean square residual; RMSEA (90% CI) = root mean square error of approximation (90% confidence interval); χ^2 = chi-square. *df* = degrees of freedom; $\Delta\chi^2$, Δdf and ΔCFI were the difference between each alternative and the configural model.

equivalent across groups. The difference in the model fit between scalar invariance and the configural model was significant: $\Delta\chi^2$ (37) = 153.868, $p < .001$ and ΔCFI = −.019; this indicated that scalar invariance was not achieved. A subsequent analysis was performed to determine which intercepts were non-invariant; it revealed only one intercept parameter that was not operating equivalently across groups (Stories Memory – Delayed Recall). If this non-invariant intercept was allowed to be freely estimated in each group (no equality constraint was imposed), the partial scalar invariance was supported: $\Delta\chi^2$ (36) = 27.901, $p = .831$ and ΔCFI = .000.

Discussion

In the last years the clinical application of neuropsychological evaluation has increased in a variety of settings. The BANC is a new comprehensive assessment instrument that taps different functions of children's neuropsychological development. It can be useful in the diagnosis of a variety of neurodevelopmental disorders and in identifying neuropsychological strengths and weaknesses. This battery can also facilitate the special education eligibility decision-making process and neurocognitive training programs.

The first main objective of the current study was to evaluate the factor structure of the BANC, to observe how the different subtests that were theoretically derived operate empirically. Only the Memory, Language and Attention/Executive Functions domains were investigated through a CFA. Consistent with the BANC's conceptualization, the three-correlated-factor model (Model 2) demonstrated an adequate overall model fit. When we analyzed local fit, the factor loadings showed primarily moderate values. Among the 20 indicators included in the three-correlated-factor model, three subtests loaded below 0.30 (Memory of Faces, Naming Speed – RAN and Tower). Similarly, Mosconi and colleagues (2008), using the standardization sample of the NEPSY, found moderate factor loadings (ranging between .31 and .69), with the Memory for Faces showing the lowest factor loading (CFA; $\lambda = 0.31$). In the Stinnett and colleagues's (2002) study the Memory for Faces from NEPSY was the second lowest factor loading in the one-factor solution (EFA; $\lambda = .26$). Interestingly, Fasfous and colleagues (2015), in their study regarding the reliability and the validity of the Battery for Neuropsychological Evaluation of Children, also reported, in general, moderate factor loadings (ranging between .18 and .82), with Planning showing the lowest factor loading (CFA; $\lambda = .18$). Future studies should explore the misspecification of these indicators in their specific factors.

The moderate factor loadings found in CFA associated with the small to medium item-total correlation coefficient of each subtest with their specific domain and the magnitude of factor reliability coefficients are consistent with the conceptualization of neuropsychological functioning as reflecting independent but related functional systems. This is particularly evident in the Attention/Executive Functions domain, which is convergent with the view of the diverse and heterogeneous nature of the executive functions (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Miyake et al., 2000; Testa, Bennett, & Ponsford, 2012). In addition, although most the subtests showed adequate temporal stability, some subtests from the Memory and Attention/Executive Functions domains revealed marginal ($r = .60 – .69$) or low ($r \leq .59$) test–retest coefficients. Thus, retest scores of these subtests should be interpreted with some caution when neuropsychologists need to repeat an evaluation. In a meta-analytic study, Calamia, Markon and Tranel (2013) also found adequate test–retest reliability coefficients ($r \geq .70$) for most neuropsychological measures, except for several memory and executive functions scores. The temporal stability of neurocognitive measures can be affected by practice effects, floor or ceiling effects, neurocognitive domain, retest interval, clinical condition, prior experiences, demographic variables, among others (see for a review: Calamia, Markon, & Tranel, 2012; Calamia et al., 2013; Duff, 2012).

Higher factor correlations between Memory, Language and Attention/Executive Functions domains were observed in the three-correlated-factor model. Mosconi and colleagues (2008) also found a higher factor correlation ($r = .85$) between Memory and Language in NEPSY. These findings are not surprising because neuropsychological tasks are often significant correlated and commonly imply other neurocognitive abilities. For example, executive functions regulate other cognitive processes, and assessing them requires other non-executive cognitive abilities (Miyake et al., 2000; van der Sluis, de Jong, & van der Leij, 2007).

Furthermore, verbal fluency is typically a measure of executive functions but also depends on language and memory abilities (Moura, Simões, et al., 2015a; Strauss et al., 2006; Whiteside et al., 2015). Phonological awareness can be significantly affected by verbal working memory (De Groot, Van den Bos, Van der Meulen, & Minnaert, 2015; Moura et al., 2017) and inattention (Martinussen, Grimbos, & Ferrari, 2014; Sims & Lonigan, 2013). Naming speed, which is often used as a language task, is also associated with processing speed (Norton & Wolf, 2012; Shanahan et al., 2006). The poor model fit obtained for the three-uncorrelated-factor model (Model 1) corroborates these findings. When the three factors were estimated without a covariance among them, the goodness-of-fit indices were clearly worse compared with the three-correlated-factor model.

A factor correlation greater than one was found between Attention and Executive Functions (Model 4). This overlap of variance is consistent with the BANC's theoretical model that combines subtests of attention and executive functions in the same domain (similar to NEPSY and NEPSY-II). Indeed, cancellation tasks are traditionally used to evaluate visual attention but are also to assess processing speed, visual scanning and discrimination, activation and inhibition of responses, among other neurocognitive abilities (Brucki & Nitrini, 2008; Lezak, Howieson, Bigler, & Tranel, 2012), whereas the Trail Making Test has been largely used to measure attention, processing speed and mental flexibility (Shanahan et al., 2006; Strauss et al., 2006). Similarly, a factor correlation greater than one was also found between Visual Memory and Attention/Executive Functions (Model 3). This finding was not initially expected, although there is empirical support for the link between executive functioning and visual memory (Beebe, Ris, Brown, & Dietrich, 2004; Duff, Schoenberg, Scott, & Adams, 2005). For example, the Rey Complex Figure is often used to investigate these relations because it assesses visual perception, planning and visual memory (Somerville, Tremont, & Stern, 2000; Watanabe et al., 2005). Thus, the four- and five-correlated-factor models were not statistically admissible. Taken together, the results from the CFA showed that the BANC's theoretically derived three-correlated-factor model provided the best fit to the data.

Because in the three-correlated-factor model (Model 2) the factors were highly correlated, a confirmatory bifactor model was estimated in order to investigate how the subtests load on their specific factor and on a general neurocognitive factor. After controlling for the general factor, the visual memory subtests (Memory of Faces, Rey Complex Figure and Corsi Block) showed very small factor loadings on the memory factor, suggesting that their variance (albeit small) is largely explained by the general neurocognitive factor. In contrast, the verbal memory subtests (Word Learning List and Stories Memory) demonstrated higher factor loadings on the specific and general factors, whereas the variance of subtests from the Language and Attention/Executive Functions factors is majority explained by the specific factor. Thus, the memory factor seems to be the less robust, probably because it includes different memory components (verbal and visual) and scores (short- and long-delay recall).

The second main objective of this study was to evaluate whether the factor structure of the BANC (three-correlated-factor model) would be equivalent across two age-groups (7–9 years and 10–15 years). The results from the multiple-group analysis supported configural invariance, which suggests that the number and pattern of factors were equivalent across groups. The full metric invariance was also established (all factor loadings were invariant), which indicates that the strength of the relation between subtests and their associated latent factors is equivalent across groups. The scalar invariance was assessed after establishing a metric invariance to evaluate whether children who have the same score on a latent factor (domain) would obtain the same score on the observed variable (subtest) regardless of their group membership (7–9 years or 10–15 years). Only the Stories Memory – Delayed Recall score contributed to the scalar non-invariance of the model (i.e., children aged 7–9 years may obtain a significantly different score on Stories Memory – Delayed Recall score relative to children aged 10–15 years with an equal score on the Memory domain). The lack of invariance might be related to the fact that the stories administered for children aged 10–15 years are more complex, resulting in different performances in the later retrieval of the stories between these two age-groups. Invoking the condition of partial measurement invariance (Byrne & van de Vijver, 2010; Vandenberg & Lance, 2000), this non-invariant parameter was released and scalar invariance was met. Thus, although some subtests have different tasks for children 7–9 years and 10–15 years, the results from the multiple-group analysis demonstrated the measurement equivalence of the BANC. These findings suggest that the same subtest score interpretation can be made across these two age-groups.

Wicherts (2016) highlighted the importance of include measurement invariance analysis in the validation of neurocognitive tests because it is crucial for the valid use in clinical, educational and professional practice. As referred by Sideridis, Tsaousis and Al-harbi (2015), unless invariance is present at least at the factor loading level (i.e., metric invariance), all subsequent between-groups comparisons (e.g., based on age, gender, neurodevelopmental disorders) may likely be suspect and invalid. To the best of our knowledge, the present study is one of the first that implement a measurement invariance analysis framework in a pediatric neuropsychological battery. Specifically, the implementation of multiple-group analysis techniques (e.g., measurement invariance, latent mean differences) may be particularly relevant to better understand and mitigate some of the limitations of applying a factor structure derived from standardization samples towards clinical populations. For example, Delis and colleagues (2003) illustrated with the California Verbal Learning Test (CVLT) that the factor structure of a neurocognitive test can change significantly depending on the clinical sample included in the analysis (Alzheimer's disease, Huntington's disease, or mixed neurological patients). Conversely, Donders and colleagues (DeJong & Donders, 2009; Donders, 2008) have found that the factor model that best fit to the data in a sample of adults with traumatic brain injury is consistent with that identified previously in a CFA of the CVLT standardization sample. A very similar finding was also observed with children with traumatic brain injury and the standardization sample of CVLT – Children's Version (Donders, 1999; Mottram & Donders, 2005).

Notwithstanding the relevance of the present study, it had some limitations that should be addressed in future studies. First, although BANC has six theoretically derived domains this study only analyzed the factor structure of the Memory, Language and Attention/Executive Functions domains. Subsequent studies should examine the factor structure of all domains and subtests. Second, it would be also particularly interesting investigate the equivalence of the factor structure between typically developing children and children with neurodevelopmental disorders.

In conclusion, the BANC is a new neuropsychological battery, with several validation studies in different clinical groups, that provides relevant information to study normal and abnormal neuropsychological development in children. This study provides evidence regarding the adequate psychometric properties of this neuropsychological battery. Specifically, these findings support the three-dimensional structure of the domains included in this study and provide evidence of the BANC's construct validity.

Funding

This work was supported by the Portuguese adaptation and standardization of neuropsychological tests: Validity and normative studies. Fundação para a Ciência e a Tecnologia (FCT).

Conflict of Interest

None declared.

References

Beebe, D. W., Ris, M. D., Brown, T. M., & Dietrich, K. N. (2004). Executive functioning and memory for the Rey-Osterreith Complex Figure task among community adolescents. *Applied Neuropsychology*, *11*, 91–98. doi:10.1207/s15324826an1102_4.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246. doi:10.1037/0033-2909.107.2.238.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*, 1641–1660. doi:10.1111/j.1467-8624.2010.01499.x.

Blunch, N. J. (2008). *Introduction to structural equation modeling using IBM SPSS Statistics and Amos*. London: Sage.

Boomsma, A. (2000). Reporting analysis of covariance structures. *Structural Equation Modeling*, *7*, 461–483. doi:10.1207/S15328007SEM0703_6.

Brucki, S. M. D., & Nitrini, R. (2008). Cancellation task in very low educated people. *Archives of Clinical Neuropsychology*, *23*, 139–147. doi:10.1016/j.acn.2007.11.003.

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In Grimm L. G., & Yarnold P. R. (Eds.), *Reading and understanding multivariate statistics* (pp. 99–136). Washington, DC: American Psychological Association.

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS Graphics: A road less traveled. *Structural Equation Modeling*, *11*, 272–300. doi:10.1207/s15328007sem1102_8.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Routledge.

Byrne, B. M., & van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*, 107–132. doi:10.1080/15305051003637306.

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: Meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, *26*, 543–570. doi:10.1080/13854046.2012.680913.

Calamia, M., Markon, K., & Tranel, D. (2013). The robust reliability of neuropsychological measures: Meta-analyses of test–retest correlations. *The Clinical Neuropsychologist*, *27*, 1077–1105. doi:10.1080/13854046.2013.809795.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, *29*, 458–472. doi:10.1037/pas0000358.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. doi:10.1207/S15328007SEM0902_5.

Coelho, S., Albuquerque, C. P., & Simões, M. R. (2013). Specific language impairment: A neuropsychological characterization. *Paidéia*, *23*, 31–41. doi:10.1590/1982-43272354201305.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16–29. doi:10.1037/1082-989X.1.1.16.

De Groot, B. J. A., Van den Bos, K. P., Van der Meulen, B. F., & Minnaert, A. E. M. G. (2015). Rapid naming and phonemic awareness in children with or without reading disabilities and/or ADHD. *Journal of Learning Disabilities*. doi:10.1177/0022219415609186.

DeJong, J., & Donders, J. (2009). A confirmatory factor analysis of the California Verbal Learning Test – Second Edition (CVLT-II) in a traumatic brain injury sample. *Assessment*, *16*, 328–336. doi:10.1177/1073191109336989.

Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment. *Journal of the International Neuropsychological Society*, *9*, 936–946. doi:10.1017/S1355617703960139.

Donders, J. (1999). Structural equation analysis of the California Verbal Learning Test – Children's Version in the standardization sample. *Developmental Neuropsychology*, *15*, 395–406. doi:10.1080/87565649909540757.

Donders, J. (2008). A confirmatory factor analysis of the California Verbal Learning Test – Second Edition (CVLT-II) in the standardization sample. *Assessment*, *15*, 123–131. doi:10.1177/1073191107310926.

Donders, J., Elzinga, B., Kuipers, D., Helder, E., & Crawford, J. R. (2013). Development of an eight-subtest short form of the WISC-IV and evaluation of its clinical utility in children with traumatic brain injury. *Child Neuropsychology*, *19*, 662–670. doi:10.1080/09297049.2012.723681.

Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives of Clinical Neuropsychology*, *27*, 248–261. doi:10.1093/arclin/acr120.

Duff, K., Schoenberg, M. R., Scott, J. G., & Adams, R. L. (2005). The relationship between executive functioning and verbal and visual learning and memory. *Archives of Clinical Neuropsychology*, *20*, 111–122. doi:10.1016/j.acn.2004.03.003.

Fasfous, A. F., Peralta-Ramirez, M. I., Pérez-Marfil, M. N., Cruz-Quintana, F., Catena-Martinez, A., & Pérez-García, M. (2015). Reliability and validity of the Arabic version of the computerized Battery for Neuropsychological Evaluation of Children (BENCI). *Child Neuropsychology*, *21*, 210–224. doi:10.1080/09297049.2014.896330.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*, 177–190. doi:10.1037/0012-1649.40.2.177.

Hu, L., & Bentler, P. M. (1999). Cutoff criterion for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6* (*1*), 1–55. doi:10.1080/10705519909540118.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*, 2017–2036. doi:10.1016/j.neuropsychologia.2006.01.010.

INE/DGOTDU. (1998). *Tipologia de áreas urbanas [The classification of urban areas]*. Lisboa: Instituto Nacional de Estatística.

Jöreskog, K. G., & Sörbom, D. G. (1989). *LISREL 7: User's reference guide*. Chicago: Scientific Software Inc.

Jöreskog, K. G., & Sörbom, D. G. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software Inc.

Khojasteh, J., & Lo, W.-J. (2015). Investigating the sensitivity of goodness-of-fit indices to detect measurement invariance in a bifactor model. *Structural Equation Modeling*, *22*, 531–541. doi:10.1080/10705511.2014.937791.

Korkman, M., Lahti-Nuuttila, P., Laasonen, M., Kemp, S. L., & Holdnack, J. (2013). Neurocognitive development in 5- to 16-year-old North American children: A cross-sectional study. *Child Neuropsychology*, *19*, 516–539. doi:10.1080/09297049.2012.705822.

Krug, D., Dean, R. S., & Anderson, J. L. (1995). Factor analysis of the Halstead-Reitan Neuropsychological Test Battery for older children. *International Journal of Neuroscience*, *83*, 131–134. doi:10.3109/00207459508986332.

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*, 59–80. doi:10.1348/026151003321164627.

Lewis M. P., Simons G. F., & Fennig C. D. (Eds.) (2015). *Ethnologue: Languages of Africa and Europe* (18th ed.). Dallas, Texas: SIL International.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York: Oxford University Press.

Livingston, R. B., Gray, R. M., Haak, R. A., & Jennings, E. (1997). Factor structure of the Halstead-Reitan Neuropsychological Test Battery for older children. *Child Neuropsychology*, *3*, 176–191. doi:10.1080/09297049708400641.

Lobley, K. J., Baddeley, A. D., & Gathercole, S. E. (2005). Phonological similarity effects in verbal complex span. *The Quarterly Journal of Experimental Psychology*, *58*, 1462–1478. doi:10.1080/02724980443000700.

Lopes, A. F., Monteiro, J. P., Fonseca, M. J., Robalo, C., & Simões, M. R. (2014). Memory functioning in children with epilepsy: Frontal lobe epilepsy, childhood absence epilepsy, and benign epilepsy with centrotemporal spikes. *Behavioural Neurology*, *2014*, 1–8. doi:10.1155/2014/218637.

Lopes, R., Simões, M. R., & Leal, A. J. R. (2014). Neuropsychological abnormalities in children with the Panayiotopoulos syndrome point to parietal lobe dysfunction. *Epilepsy & Behavior*, *31*, 50–55. doi:10.1016/j.yebeh.2013.11.013.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320–341. doi:10.1207/s15328007sem1103_2.

Martinussen, R., Grimbos, T., & Ferrari, J. L. S. (2014). Word-level reading achievement and behavioral inattention: Exploring their overlap and relations with naming speed and phonemic awareness in a community sample of children. *Archives of Clinical Neuropsychology*, *29*, 680–690. doi:10.1093/arclin/acu040.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568–592.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. doi:10.1006/cogp.1999.0734.

Mosconi, M., Nelson, L., & Hooper, S. R. (2008). Confirmatory factor analysis of the NEPSY for younger and older school-age children. *Psychological Reports*, *102*, 861–866. doi:10.2466/pr0.102.3.861-866.

Mottram, L., & Donders, J. (2005). Construct validity of the California Verbal Learning Test – Children's Version (CVLT-C) after pediatric traumatic brain injury. *Psychological Assessment*, *17*, 212–217. doi:10.1037/1040-3590.17.2.212.

Moura, O., Moreno, J., Pereira, M., & Simões, M. R. (2015). Developmental dyslexia and phonological processing in European Portuguese orthography. *Dyslexia*, *21*, 60–79. doi:10.1002/dys.1489.

Moura, O., Pereira, M., Alfaiate, C., Fernandes, E., Fernandes, B., Nogueira, S., et al. (2017). Neurocognitive functioning in children with developmental dyslexia and attention-deficit/hyperactivity disorder: Multiple deficits and diagnostic accuracy. *Journal of Clinical and Experimental Neuropsychology*, *39*, 296–312. doi:10.1080/13803395.2016.1225007.

Moura, O., Simões, M. R., & Pereira, M. (2015a). Executive functioning in children with developmental dyslexia. *The Clinical Neuropsychologist*, *28*, 20–41. doi:10.1080/13854046.2014.964326.

Moura, O., Simões, M. R., & Pereira, M. (2015b). Working memory in Portuguese children with developmental dyslexia. *Applied Neuropsychology: Child*, *4*, 237–248. doi:10.1080/21622965.2014.885389.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, *63*, 427–452. doi:10.1146/annurev-psych-120710-100431.

Pickering, S. J. (2004). Verbal memory in the learning of literacy. In Turner M., & Rack J. (Eds.), *The study of dyslexia* (pp. 131–156). New York: Kluwer Academic Publishers.

Reynolds, C. R., & Mason, B. A. (2009). Measurement and statistical problems in neuropsychological assessment of children. In Reynolds C. R., & Fletcher-Janzen E. (Eds.), *Handbook of clinical child neuropsychology* (3rd ed., pp. 203–230). New York: Springer.

Ross, S. A., Allen, D. N., & Goldstein, G. (2014). Factor structure of the Halstead-Reitan Neuropsychological Battery for Children: A brief report supplement. *Applied Neuropsychology: Child*, *3* (*1*), 1–9. doi:10.1080/21622965.2012.695882.

Sá, D. S. F., Albuquerque, C. P., & Simões, M. R. (2008). Avaliação neuropsicológica da perturbação de oposição e desafio [Neuropsychological assessment of the oppositional defiant disorder]. *Psicologia, Saúde & Doenças*, *9*, 299–317.

Santos, L. M. G. (2006). *Consequências neuropsicológicas, comportamentais e sócio-emocionais dos traumatismos crânio-encefálicos na infância e adolescência: Um estudo exploratório [Neuropsychological consequences, behavioral and socio-emotional of traumatic brain injury in childhood and adolescence: An exploratory study*]. Master Degree in Psychological Assessment, Faculty of Psychology and Educational Sciences. University of Coimbra.

Shanahan, M. A., Pennington, B. F., Yerys, B. E., Scott, A., Boada, R., Willcutt, E. G., et al. (2006). Processing speed deficits in attention deficit/hyperactivity disorder and reading disability. *Journal of Abnormal Child Psychology*, *34*, 584–601. doi:10.1007/s10802-006-9037-8.

Shuttleworth-Edwards, A. B. (2016). Generally representative is representative of none: Commentary on the pitfalls of IQ test standardization in multicultural settings. *The Clinical Neuropsychologist*, *30*, 975–998. doi:10.1080/13854046.2016.1204011.

Sideridis, G. D., Tsaousis, I., & Al-harbi, K. A. (2015). Multi-population invariance with dichotomous measures: Combining multi-group and MIMIC methodologies in evaluating the General Aptitude Test in the Arabic language. *Journal of Psychoeducational Assessment*, *33*, 568–584. doi:10.1177/0734282914567871.

Simões, M. R., Albuquerque, C. P., Pinho, M. S., Vilar, M., Pereira, M., Lopes, A. F., et al. (2016). *Bateria de Avaliação Neuropsicológica de Coimbra (BANC) [Coimbra Neuropsychological Assessment Battery]*. Lisboa: CEGOC-TEA.

Sims, D. M., & Lonigan, C. J. (2013). Inattention, hyperactivity, and emergent literacy: Different facets of inattention relate uniquely to preschoolers' reading-related skills. *Journal of Clinical Child and Adolescent Psychology*, *42*, 208–219. doi:10.1080/15374416.2012.738453.

Somerville, J., Tremont, G., & Stern, R. A. (2000). The Boston Qualitative Scoring System as a measure of executive functioning in Rey-Osterrieth Complex Figure performance. *Journal of Clinical and Experimental Neuropsychology*, *22*, 613–621. doi:10.1076/1380-3395(200010)22:5;1-9;FT613.

Stinnett, T. A., Oehler-Stinnett, J., Fuqua, D. R., & Palmer, L. S. (2002). Examination of the underlying structure of the NEPSY: A Developmental Neuropsychological Assessment. *Journal of Psychoeducational Assessment*, *20*, 66–82. doi:10.1177/073428290202000105.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Testa, R., Bennett, P., & Ponsford, J. (2012). Factor analysis of nineteen executive function tests in a healthy adult population. *Archives of Clinical Neuropsychology*, *27*, 213–224. doi:10.1093/arclin/acr112.

Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faísca, L., Reis, A., & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology*, *102*, 827–842. doi:10.1037/a0019465.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*, 427–449. doi:10.1016/j.intell.2006.09.001.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70. doi:10.1177/109442810031002.

Watanabe, K., Ogino, T., Nakano, K., Hattori, J., Kado, Y., Sanada, S., & Ohtsuka, Y. (2005). The Rey–Osterrieth Complex Figure as a measure of executive function in childhood. *Brain and Development*, *27*, 564–569. doi:10.1016/j.braindev.2005.02.007.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WISC-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psychoeducational Assessment*, *31*, 114–131. doi:10.1177/0734282913478032.

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2015). Verbal fluency: Language or executive function measure? *Applied Neuropsychology: Adult*, *23* (*1*), 1–6. doi:10.1080/23279095.2015.1004574.

Wicherts, J. M. (2016). The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist*, *30*, 1006–1016. doi:10.1080/13854046.2016.1205136.