

# An Empirical Comparison of Competing Factor Structures for the Repeatable Battery for the Assessment of Neuropsychological Status: A Project FRONTIER Study<sup>†</sup>

Nicole D. Torrence<sup>1,3</sup>, Samantha E. John<sup>1</sup>, Brandon E. Gavett<sup>1,\*</sup>, Sid E. O'Bryant<sup>2</sup>

<sup>1</sup>Department of Psychology, University of Colorado Colorado Springs, Colorado Springs, CO, USA

<sup>2</sup>Department of Internal Medicine and Institute for Aging & Alzheimer's Disease Research, University of North Texas Health Science Center, Fort Worth, TX, USA

\*Corresponding author at: Department of Psychology, University of Colorado Colorado Springs, 1420 Austin Bluffs Parkway, Colorado Springs, CO 80918, USA.

Tel.: +1-719-255-4135; fax: +1-719-255-4166.

E-mail address: bgavett@uccs.edu (B.E. Gavett).

Accepted 2 September 2015

## Abstract

The original factor structure of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) has received little empirical support, but at least eight alternative factor structures have been identified in the literature. The current study used confirmatory factor analysis to compare the original RBANS model with eight alternatives, which were adjusted to include a general factor. Participant data were obtained from Project FRONTIER, an epidemiological study of rural health, and comprised 341 adults (229 women, 112 men) with mean age of 61.2 years ( $SD = 12.1$ ) and mean education of 12.4 years ( $SD = 3.3$ ). A bifactor version of the model proposed by Duff and colleagues provided the best fit to the data ( $CFI = 0.98$ ; root-mean-squared error of approximation = 0.07), but required further modification to produce appropriate factor loadings. The results support the inclusion of a general factor and provide partial replication of the Duff and colleagues RBANS model.

**Keywords:** Validity; Factor analysis; Statistical model; Cognition; Neuropsychology

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) is an individually administered, theoretically informed assessment instrument designed to be a stand-alone, brief, and repeatable tool for assessing cognitive performance (Randolph, 1998). Made up of 12 tasks, it yields a global measure of cognitive functioning as well as scores for five specific domains: Attention, Language, Visuospatial/Constructional, Immediate Memory, and Delayed Memory. The original factor structure proposed by Randolph (1998) has received little empirical support, and as a result, alternative factor structures have been proposed (Carrlozzi, Horner, Yang, & Tilley, 2008; Duff et al., 2006; Garcia, Leahy, Corradi, & Forchetti, 2008; King, Bailie, Kinney, & Nitch, 2012; Schmitt et al., 2010; Wilde, 2006; Yang, Garrett-Mayer, Schneider, Gollomp, & Tilley, 2009). However, unlike the original RBANS model (Randolph, 1998), most of the alternative models were not designed to incorporate a general cognitive factor (as in the RBANS Total Scale). Although some studies have examined the fit of a single-factor model (e.g., Duff et al., 2006), most of this previous research has not closely examined both general (e.g., global cognition) and specific (e.g., language) factors within the same model of the RBANS. The one exception, however, was a study by King and colleagues (2012), which found strong support for the inclusion of a general cognitive factor. That study will be discussed in more detail below.

One study to date, which used confirmatory factor analysis (CFA) methods, has supported a model that is closely related to the original factor structure of the RBANS (Cheng et al., 2011). Because that study only included the five RBANS subfactors, and not a

<sup>†</sup> The results in this paper have been presented as a poster at the 2012 conference of the National Academy of Neuropsychology in Nashville, TN. The scoring program for the RBANS model described in this paper can be accessed at <http://begavett.shinyapps.io/RBANS>.

<sup>3</sup> Present address of NT: VA Puget Sound Health Care System, American Lake, USA.

general factor, it did not provide an exact test of the original RBANS factor structure. In addition, that study was performed using a Chinese translation of the RBANS, which introduced other confounds that might potentially make it more difficult to draw inferences about the original English version of the test.

Although [Cheng and colleagues \(2011\)](#) used CFA methods, the majority of studies in this area have used either exploratory factor analysis (EFA) or principal components analysis (PCA) to study the factor structure of the RBANS. [Carrlozzi and colleagues \(2008\)](#) used EFA with orthogonal rotation, followed by CFA, to determine the most appropriate factor structure in a sample of 175 male veterans. A two-factor structure with Memory and Visuospatial Skills as latent variables was the model that emerged. The [Carrlozzi and colleagues \(2008\)](#) model fit best when Digit Span was removed due to ceiling effects and poor factor loadings, and with cross-loadings of Coding and Figure Recall on the two identified factors.

In a large and diverse sample of 824 healthy older adults, [Duff and colleagues \(2006\)](#) utilized EFA with both oblique and orthogonal rotation; both methods produced similar results. Consistent with the results of [Carrlozzi and colleagues \(2008\)](#), the best fitting model reported by [Duff and colleagues \(2006\)](#) included two factors: verbal memory and visual processing; these results were based on data that excluded the Digit Span and Semantic Fluency subtests. It should be noted that [Duff and colleagues \(2006\)](#) explored the fit of a single-factor model, yet found this model to fit poorly.

In contrast to the other alternative models, four studies retained all of the RBANS subtests in the analyses. [Garcia and colleagues \(2008\)](#) explored the component structure of the RBANS using PCA with oblique rotation in a sample of 351 older adults with memory problems and identified a three component structure of the RBANS: Verbal Processing, Visuomotor Processing, and Memory. In a homogeneous sample of stroke patients, [Wilde \(2006\)](#) used PCA with orthogonal rotation and reported that, when all items were included in the model, the best underlying structure included two components: Language/Verbal Memory and Visuospatial/Visual Memory. Another recent study that explored the factor structure of the RBANS was completed using PCA in a large sample of 636 adults referred for a dementia evaluation. The results yielded a two-component model, including Visuospatial/Attention and Memory/Language ([Schmitt et al., 2010](#)).

As mentioned above, a recent study by [King and colleagues \(2012\)](#) is noteworthy, because it used principal factor analysis to find strong support for a general factor, as well as two subfactors (memory and a heterogeneous factor combining visuospatial skills and processing speed). Because these results were obtained using a sample of 167 inpatients diagnosed with schizophrenia or schizoaffective disorder, it is unclear whether they will generalize beyond this highly unique sample.

One publication to date has directly compared competing RBANS models, but performed these comparisons in a highly selected sample: older adults with early untreated Parkinson's disease. [Yang and colleagues \(2009\)](#) used CFA to assess the factor structures proposed by [Randolph \(1998\)](#), [Carrlozzi and colleagues \(2008\)](#), [Garcia and colleagues \(2008\)](#), [Duff and colleagues \(2006\)](#), and [Wilde \(2006\)](#) within their sample ( $n = 383$ ). The results presented by [Yang and colleagues \(2009\)](#) suggested that none of the competing models provided a good fit of their data, and so they utilized EFA with orthogonal rotation and stringent item correlation coefficients ( $> .40$ ) to identify a model of best fit. The best fitting model had a two-factor structure (factor labels not provided) that included only 6 of the 12 RBANS tasks (all of which involved episodic memory), but that accounted for 100% of the variance.

This collection of research focused on the factor structure of the RBANS has revealed several key ideas important for future investigations. First, because every study has produced its own unique model of best fit, the reported factor structures may have been dependent upon sample characteristics, especially in studies where EFA or PCA has been used, as these methods are empirical, rather than theory-driven. In addition, some studies have retained all of the subtests, whereas others were forced to exclude subtests (e.g., Digit Span) in order to arrive at an appropriate solution. Finally, it should be noted that only one of the aforementioned studies came close to replicating the original factor structure of the RBANS: [Cheng and colleagues \(2011\)](#) found support for a five-factor model without a general factor. In fact, none of the other studies found evidence for more than three specific cognitive domains measured by the RBANS, compared with the five specific domains proposed by [Randolph \(1998\)](#). This failure to replicate should be interpreted in light of the fact that only one of the previous studies ([King et al., 2012](#)) incorporated both general and specific factors together in one model.

General factors are common in the measurement of cognition. For instance, the Wechsler Adult Intelligence Scales ([Wechsler, 2008](#)), Wechsler Memory Scales ([Wechsler, 2009](#)), Neuropsychological Assessment Battery ([Stern & White, 2003](#)), and Woodcock–Johnson ([McGrew, 2009](#); [McGrew, Schrank, & Woodcock, 2007](#)) batteries, for example, all provide a general cognitive ability estimate in addition to more specific subfactors. Therefore, an important hypothesis to be tested with CFA relates to whether or not the inclusion of a general factor in models of the RBANS contributes to improvements in model fit.

Bifactor and second-order CFA designs are commonly used to incorporate general and specific factors in the same model ([Reise, Moore, & Haviland, 2010](#)). The bifactor model directly relates test scores to specific subfactors (e.g., Memory, Language) as well as to a general factor permeating the model ([Gavett, Crane, & Dams-O'Connor, 2013](#); [Gibbons & Hedeker, 1992](#); [Gibbons et al., 2007](#); [Reise, Morizot, & Hays, 2007](#)). Second-order models do not propose a direct relationship between the general factor and the observed test scores; rather these models hypothesize that the test results provide an indirect measure

of the general factor, with specific subfactors (e.g., Memory, Language) directly related to the observed test scores and subsumed under the general factor (Chen, West, & Sousa, 2006). For the RBANS, the general factor is the construct represented by the “Total Scale” and may provide a measure of overall cognitive ability (Randolph, 1998).

As reviewed above, most studies investigating the factor structure of the RBANS have identified unique models; the unique findings reported by each study have not moved the field toward a consensus for how to interpret RBANS scores. Only Yang and colleagues (2009) have used a model comparison approach, but failed to find support for an existing model. The goal of the current study is to build upon the ideas of Yang and colleagues (2009) by directly comparing the nine different RBANS models against one another. Although we have no reason to make any *a priori* hypotheses about which model will provide the best fit to the data, we do hypothesize that the inclusion of one general factor and one or more specific subfactors in the same model will provide better fit than models that do not include a general factor.

## Method

### Participants

For this study, we used archival data from Project FRONTIER (Facing Rural Obstacles to healthcare Now Through Intervention, Education, and Research), an epidemiological study of rural aging in the state of Texas. Participants were aged 40 and older at time of participation and resided in the counties of Cochran, Bailey, and Parmer; additional information about recruitment and selection of participants can be found in O’Bryant, Hall, and colleagues (2010) and O’Bryant, Lacritz, and colleagues (2010). We began with a sample of 585 participants from Project FRONTIER and excluded those who were not tested in English or who produced invalid results (see procedure below). The application of these exclusion criteria led to a sample of 341 participants on which all analyses are based. This sample included 229 women and 112 men, with mean years of age and education of 61.2 ( $SD = 12.1$ ) and 12.4 ( $SD = 3.3$ ), respectively. There were 281 participants (82.4%) who identified their race as White, 13 (3.8%) as Black or African American, and one as another unspecified race. Data on race were unavailable for 46 participants (13.5%). There were 97 participants (28.4%) who identified their ethnic background as Hispanic, Latino, or Spanish in origin; the remaining 244 participants (71.6%) reported a non-Hispanic, Latino, or Spanish origin. Based on the evaluation performed as part of Project FRONTIER, participants were assigned a consensus diagnosis according to commonly accepted criteria for neurodegenerative dementias (see Menon, Hall, Hobson, Johnson, & O’Bryant, 2012). The current sample was made up of cognitively healthy individuals, both with ( $n = 256$ ) and without subjective cognitive impairment ( $n = 22$ ) as well as those diagnosed with mild cognitive impairment ( $n = 57$ ) and dementia ( $n = 4$ ). The overall sample achieved an average score on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975) of 28.2 ( $SD = 2.1$ , interval = 19–30).

### Procedure

For each participant, the RBANS was individually administered at one time point, according to standardized instructions. Standard scoring procedures were implemented as described in the RBANS manual (Randolph, 1998). At consensus review, consistencies and inconsistencies within cognitive testing (e.g., digit span, list recognition) were reviewed to make a determination regarding the validity of the results. No formal symptom validity testing was administered as part of the protocol and validity of test results was based on overall judgment by neuropsychologists reviewing the individual test scores and items.

### Statistical Analysis

We used Mplus version 6.11 (Muthén & Muthén, 1998–2010) to conduct confirmatory factor analyses comparing nine competing models of the RBANS factor structure: the original model (Randolph, 1998) and eight alternative models (Carrlozzi et al., 2008; Cheng et al., 2011; Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010; Wilde, 2006; Yang et al., 2009). Only two of these models as originally proposed were made up of both a general factor and specific subfactors (King et al., 2012; Randolph, 1998). The other seven models were based on first-order factor structures that we modified to include a global index of cognition, allowing each to be examined with both bifactor and second-order factor structures. With the exception of the original model (Randolph, 1998) and the King and colleagues (2012) model, we evaluated the fit of each model using three variations: first-order, bifactor, and second-order. The Randolph (1998) and King and colleagues (2012) models were evaluated as bi-factor and second-order models, but not as first-order models. For all models, we fixed the variance of each latent variable to 1 in order to ensure compatibility with traditional scoring methods (each RBANS factor score has the same mean and  $SD$ ). We used a maximum likelihood estimator that is robust to nonnormality (the MLR estimator in Mplus) for all analyses and assessed model fit

using the root-mean-squared error of approximation (values  $<0.06$  reflect good fit), comparative fit index (CFI) and Tucker–Lewis index (TLI; values  $\geq 0.95$  reflect good fit), and standardized root-mean-squared residual (values  $\leq 0.08$  reflect good fit; Bentler, 1990; Browne & Cudeck, 1993; Hu & Bentler, 1999). Although we report the  $\chi^2$  statistic, this was not strongly considered in the comparisons, as it is influenced by large sample size.

## Results

We first used CFA to examine the fit of the nine competing models, as they were originally proposed. None of these models were supported, as indicated by poor fit statistics (Table 1). We then compared second-order and bifactor structures for each of the competing models with the addition of a general factor to each of the seven first-order models. The bifactor models were consistently superior to the second-order models (data not shown), and therefore, all results reported exclude the second-order models. Overall, the two best fitting models were the bifactor versions of the Duff and colleagues (2006) and Yang and colleagues (2009) models (Table 2). Because the Yang and colleagues (2009) model was only based on 6 of 12 RBANS tests, all of which involve episodic memory, the utility of that model was thought to be less than the utility of the Duff and colleagues (2006) model, which included 9 of 12 RBANS tests that covered a broader range of cognitive abilities. Thus, considering the relative equivalence in fit between these two models, we chose to continue with additional analyses based on the Duff and colleagues (2006) model for practical purposes.

Although the fit indices for the bifactor variant of the Duff and colleagues (2006) model were superior to those for all models other than Yang and colleagues (2009), the Duff model nevertheless produced inappropriate (i.e., negative) parameter estimates for Story Memory and Story Recall within the Verbal Memory factor (Table 3). One possible reason for these negative parameter estimates may have been due to the fact that the Duff and colleagues (2006) model did not account for method effects (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). In particular, the use of the same list of words within the List Learning and List Recall tasks, the same story within the Story Memory and Story Recall tasks, and the same figure within the Figure Copy and Figure Recall tasks may produce covariance in test scores that is not related to the underlying abilities being measured. In addition, the bifactor version of the Duff and colleagues (2006) model results yielded two non-significant parameter estimates: List Learning ( $\beta = 0.14$ ,  $p =$

**Table 1.** Fit statistics for the original versions of the Repeatable Battery for the Assessment of Neuropsychological Status models

Model	$\chi^2$	<i>df</i>	<i>p</i>	CFI	TLI	RMSEA (90% CI)	SRMR
Cheng model (first order) <sup>a</sup>	350.67	44	<.001	0.80	0.70	0.14 (0.13–0.16)	0.07
Carlozzi model (first order)	324.81	42	<.001	0.80	0.74	0.14 (0.13–0.16)	0.16
Randolph model (second order)	386.20	50	<.001	0.78	0.71	0.14 (0.13–0.15)	0.08
Wilde model (first order)	500.95	54	<.001	0.71	0.65	0.16 (0.14–0.17)	0.21
Schmitt model (first order)	590.47	54	<.001	0.65	0.58	0.17 (0.16–0.18)	0.21
Duff model (first order)	383.80	27	<.001	0.71	0.61	0.20 (0.18–0.22)	0.22
Garcia model (first order)	829.20	54	<.001	0.50	0.39	0.21 (0.19–0.22)	0.27
Yang model (first order)	221.04	9	<.001	0.74	0.57	0.26 (0.23–0.29)	0.25

Notes: CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation; CI = confidence intervals; SRMR = standardized root-mean-square residual.

<sup>a</sup>Not positive definite latent variable covariance matrix due to a correlation of 1.09 between Immediate and Delayed Memory factors.

**Table 2.** Fit statistics for bifactor versions of the Repeatable Battery for the Assessment of Neuropsychological Status models

Title	$\chi^2$	<i>df</i>	<i>p</i>	CFI	TLI	RMSEA (90% CI)	SRMR
Modified Duff model (bifactor)	31.75	13	.003	0.98	0.97	0.07 (0.04–0.09)	0.03
Duff model (bifactor)	45.69	18	<.001	0.98	0.95	0.07 (0.04–0.09)	0.04
Yang model (bifactor)	10.03	3	.018	0.99	0.96	0.08 (0.03–0.14)	0.02
Carlozzi model (bifactor)	95.15	31	<.001	0.96	0.92	0.08 (0.06–0.10)	0.04
Wilde model (bifactor)	162.98	42	<.001	0.92	0.88	0.09 (0.08–0.11)	0.06
King model (bifactor)	170.94	43	<.001	0.92	0.87	0.09 (0.08–0.11)	0.06
Schmitt model (bifactor)	254.91	42	<.001	0.86	0.78	0.12 (0.11–0.14)	0.07
Garcia model (bifactor)	287.34	42	<.001	0.84	0.75	0.13 (0.12–0.15)	0.08
Randolph model (bifactor) <sup>a</sup>				Model did not converge			

Notes: CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation; CI = confidence intervals; SRMR = standardized root-mean-square residual.

<sup>a</sup>The bifactor and second-order structures make the Randolph and Cheng models equivalent; therefore, the Cheng model is not reported here.

**Table 3.** Standardized parameter estimates and *p*-values for the original and bifactor modifications of the [Duff and colleagues \(2006\)](#) model

Factor	Measure	Duff and colleagues (2006) original		Bifactor Modification 1 <sup>a</sup>		Bifactor Modification 2 <sup>b</sup>	
		$\beta$	<i>p</i>	$\beta$	<i>p</i>	$\beta$	<i>p</i>
Visual Processing	Coding	0.57	<.001	0.21	<.001	0.21	.01
	Figure copy	0.72	<.001	0.69	<.001	0.43	<.001
	Figure recall	0.82	<.001	0.60	<.001	0.37	<.001
	Line orientation	0.46	<.001	0.31	<.001	0.45	.001
Verbal Memory	List learning	0.75	<.001	0.14	.09	N/A	N/A
	List recall	0.74	<.001	0.34	.01	N/A	N/A
	List recognition	0.00	.98	0.21	.003	N/A	N/A
	Story memory	0.83	<.001	−0.45	<.001	N/A	N/A
	Story recall	0.88	<.001	−0.35	<.001	N/A	N/A
Global	Coding	N/A	N/A	0.66	<.001	0.72	<.001
	Figure copy	N/A	N/A	0.35	<.001	0.38	<.001
	Figure recall	N/A	N/A	0.53	<.001	0.56	<.001
	List learning	N/A	N/A	0.82	<.001	0.81	<.001
	List recall	N/A	N/A	0.88	<.001	0.81	<.001
	List recognition	N/A	N/A	0.04	.56	N/A	N/A
	Line orientation	N/A	N/A	0.31	<.001	0.33	<.001
	Story memory	N/A	N/A	0.76	<.001	0.67	<.001
	Story recall	N/A	N/A	0.83	<.001	0.77	<.001

Notes: <sup>a</sup>Does not account for method effects and includes List Recognition.

<sup>b</sup>Accounts for method effects between lists and stories and does not include List Recognition.

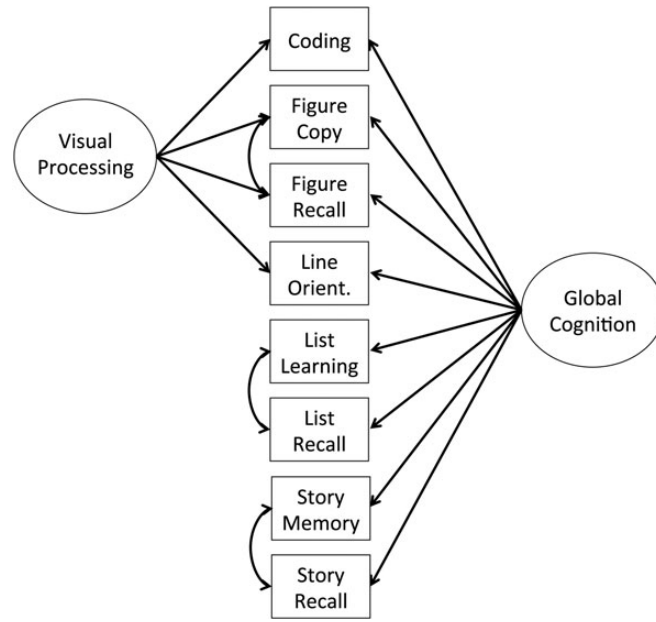
.09) within the Verbal Memory factor and List Recognition ( $\beta = 0.04$ ,  $p = .56$ ) within the Global factor. Therefore, we applied a further modification to the [Duff and colleagues \(2006\)](#) bifactor model to account for method effects, at the same time removing List Recognition from the model. Within the Verbal Memory factor, we found that accounting for method effects by estimating residual correlations between the immediate and delayed trials of the two verbal memory tasks improved model fit and yielded appropriate parameter estimates. Consequently, the model fit better without a Verbal Memory factor than with this factor present (see Fig. 1 for a depiction of the best fitting model). This modification of the bifactor [Duff and colleagues \(2006\)](#) model included one general factor and one specific factor—Visual Processing—measured by Coding, Figure Copy, Figure Recall, and Line Orientation.

Based on the results of the bifactor modification of the [Duff and colleagues \(2006\)](#) model, we generated factor scores for Global Cognition and Visual Processing for each participant, rescaled to match the mean and *SD* of the RBANS scores. These factor scores were compared with the original RBANS Index Scores in order to compare the agreement between the two methods for measuring global cognitive ability and visuospatial ability. These comparisons are depicted graphically in Fig. 2. In the current sample, the two global index scores were compared with another measure of global cognition, the Clinical Dementia Rating Sum of Boxes (CDR-SB; [O'Bryant, Hall, et al., 2010](#); [O'Bryant, Lacritz, et al., 2010](#)). Although the alternative model global factor scores were more strongly related to CDR-SB scores (Spearman's  $\rho = -0.47$ ) than the RBANS Total Index scores (Spearman's  $\rho = -0.45$ ), the difference in correlation coefficients was negligible. This pattern was the same when using Trail Making Test part B ([Reitan & Wolfson, 1985](#)) as the criterion measure (Spearman's  $\rho = -0.61$  for the alternative global model factor scores and  $-0.59$  for the RBANS Total Index scores). Similarly, the two visuospatial test scores were compared with another visuospatial tests, the copy condition (CLOX2) of the Executive Clock Drawing Test ([Royall, Cordes, & Polk, 1995](#)), and similar findings emerged; the alternative model visual processing factor scores were marginally more related to CLOX2 scores (Spearman's  $\rho = 0.34$ ) than the RBANS Visuospatial/Constructional Index score (Spearman's  $\rho = 0.32$ ). Finally, when comparing individuals in our sample with and without documented cognitive impairment, the classification accuracy of the alternative model global factor scores [AUC = 0.875, 95% CI (0.825, 0.925)] was negligibly higher than the RBANS Global Index scores [AUC = 0.867, 95% CI (0.817, 0.917)].

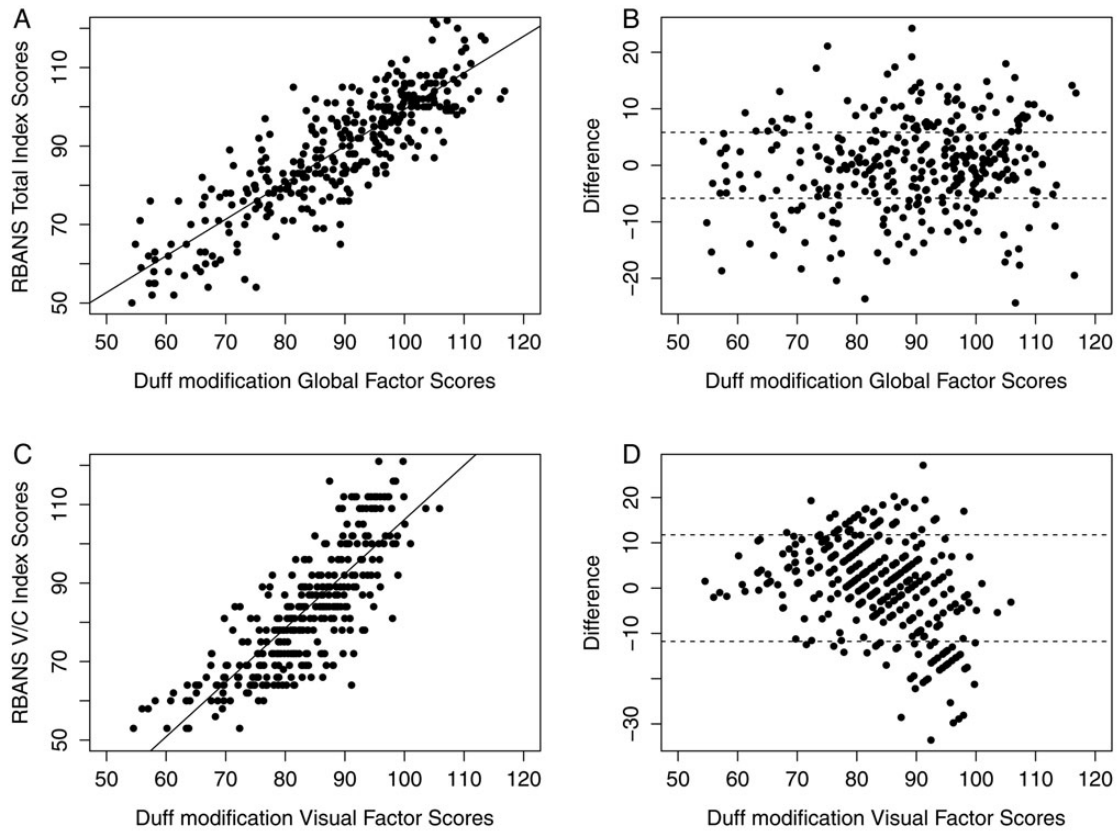
## Discussion

Previous research has failed to reach a consensus about the factor structure underlying performance on the RBANS. Although at least eight different studies have attempted to validate the original structure proposed in the RBANS manual ([Randolph 1998](#)), only one ([Cheng et al., 2011](#)) has come close to supporting this model. The model supported by [Cheng and colleagues \(2011\)](#) omitted a general cognitive factor, which is almost always relevant in comprehensive cognitive batteries (e.g., [McGrew, 2009](#);





**Fig. 1.** The best fitting model was a bifactor version of the [Duff and colleagues \(2006\)](#) model, modified to account for method effects (curved double-headed arrows) and to eliminate the List Recognition test and Verbal Processing factor from the model.



**Fig. 2.** Agreement between factor scores estimated from the alternative model (the bifactor modification of the [Duff et al., 2006](#) model) and the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) Index Scores for global cognition (A and B) and visuospatial functioning (C and D). Panels on the left (A and C) show scatterplots between the alternative model factor scores on x axis and RBANS Index scores on y axis, with an ordinary least squares line of best fit drawn through the data. Panels on the right (B and D) show how differences between alternative model factor scores and RBANS Index scores (y axis) vary according to the alternative model factor scores (x axis). Higher difference scores are produced when alternative model factor scores are larger than RBANS Index scores. Dashed lines represent  $\pm 1$  standard error of the estimate.

McGrew et al., 2007; Stern & White, 2003; Wechsler, 2008, 2009), and which was found to improve the fit of the RBANS in the previous research (King et al., 2012). In order to compare the eight RBANS factor structure variants with one another and with the original RBANS model, we used a series of hypothesis-driven CFAs to examine model fit to our data. Of the nine original models examined, all demonstrated suboptimal fit. Because these RBANS models were not supported within our sample, we modified them with the goal of incorporating a general factor into the models by applying bifactor and second-order structures. Overall, the bifactor models demonstrated superior fit to the second-order models and a bifactor modification of the Duff and colleagues (2006) model produced the best combination of fit statistics and content coverage.

Despite the fact that a bifactor model that added a general factor to the original Duff and colleagues (2006) model fit the data well, we were able to improve upon that model by replacing the Verbal Memory factor with test-specific method effects accounting for residual correlations and by eliminating the List Recognition test from the model. Duff and colleagues (2009) provide normative data for their Verbal Memory and Visual Processing factor scores. However, the current results suggest interpreting these normative data and factor scores with caution. The Duff and colleagues (2006) model and normative data (Duff et al., 2009) were developed without accounting for a general factor, which, based on our results and those reported by King and colleagues (2012), adds substantial improvement in fit to all first-order models, including the Duff and colleagues (2006) model. As can be seen in Table 3, the factor loadings for modifications of the Duff and colleagues (2006) model varied considerably depending on whether or not a general factor was used to explain the variance across all test scores. In addition, our results suggest that, after accounting for global cognitive ability and method effects within lists and stories, a Verbal Memory factor no longer provided the best fit to the data.

Our approach to studying the RBANS is unique in several ways. Although we are not the first to use CFA to test alternative models that have been proposed in the literature (see Yang et al., 2009), our study differs from Yang and colleagues (2009), because they were unable to find support for any competing RBANS models within their sample of individuals with Parkinson's disease. Consistent with the Yang and colleagues (2009) study, we failed to find support for any of the competing RBANS models as originally proposed. In contrast to the Yang and colleagues (2009) study, we tested the hypothesis that the addition of a general factor would improve model fit and, in doing so, found evidence supporting one of the competing models. It remains to be seen whether support for the bifactor modification of Duff and colleagues (2006) can be replicated in homogeneous clinical samples such as those studied by Yang and colleagues (2009) or King and colleagues (2012).

Our results have the potential to rectify conflicting results in the literature, which may have stemmed from exploratory methods and sampling artifacts. Because we used a large and diverse sample of rural dwelling adults, these results may be more generalizable than studies conducted with highly selected study participants. The current study provides empirical support for the superiority of a bifactor model, which provides several advantages: bifactor models account for general and specific cognitive abilities together and can be scored so that they possess linear measurement properties and interval scaling. Linear measurement properties reflect a uniform relationship between cognitive ability and test scores (Mungas & Reed, 2000) and are highly desirable for longitudinal applications, which is an obvious focus of the RBANS. The traditional test scores produced by the RBANS are scaled ordinally, which can be problematic in meeting the assumptions of many statistical analysis methods and for tracking changes in cognition over time.

The current results are also important for clinical purposes in that they suggest RBANS test scores cannot reliably distinguish between specific cognitive abilities. Our data suggest that only global cognitive functioning and visual processing can be accurately measured by the RBANS. But as suggested by King and colleagues (2012), the large number of memory tests produced by the RBANS means that the general factor may largely be a measure of memory ability. Clinicians or researchers who interpret the RBANS according to its original model (with a global index score along with five specific index scores) are likely to be misled when the test is used to estimate latent abilities other than global cognition and visual processing (e.g., language and attention). In addition, the results do not support the use of RBANS test scores to distinguish between the Immediate and Delayed components of episodic memory. The recent RBANS normative update, which provides normative data for each individual test, may be more clinically useful than interpreting the RBANS according to its original five factor scores.

As part of our analyses, we compared the global and visual processing factor scores generated by the bifactor modification of the Duff and colleagues (2006) model to the traditional RBANS index scores (Fig. 2). Although the general trends in the data are similar between the two approaches, there are also numerous data points that are highly discrepant. The dashed horizontal lines in Fig. 2B and D represent  $\pm 1$  standard error of the estimate for the factor scores; data points beyond these markers represent score differences that are outside the margin for random error in the factor score estimates. For some individuals, the factor scores generated from the bifactor modification of the Duff model will produce ability estimates that deviate considerably from the RBANS index scores when obtained using the traditional approach to scoring the test. The criterion validity appears to be roughly equivalent when comparing the bifactor modification of the Duff and colleagues (2006) model and the traditional RBANS approach to generating index scores, when using the CDR-SB, Trail Making Test Part B, the Executive Clock Drawing Test, and consensus diagnosis of cognitive impairment as the criterion standards. Very slight statistical advantages

were consistently observed in favor of the alternative model factor scores, but these differences are unlikely to be clinically meaningful. Because of the longitudinal advantages of using bifactor scaling instead of the traditional RBANS approach to scoring, future studies should also make criterion validity comparisons between the two methods in the context of serial assessments.

The current study should be interpreted in light of several limitations. [Delis, Jacobson, Bondi, Hamilton, and Salmon \(2003\)](#) argued that replication in multiple homogeneous clinical samples is the preferred method for validating the underlying factor structure of neuropsychological tests, an approach that we did not utilize. Although our participants were diverse in terms of the nature of their cognitive functioning, most individuals diagnosed with cognitive impairment demonstrated relatively mild deficits. Therefore, these findings should be confirmed by replication in multiple homogeneous clinical samples. It should also be noted that there are two methods that can be used for scoring the RBANS Figure Copy and Figure Recall subtests. In the current study, we used the scoring criteria as described in the RBANS manual, which is different than the scoring criteria ([Duff et al., 2007](#)) that were used by [Duff and colleagues \(2006\)](#) to establish their RBANS factor structure. Although it is uncertain what differences, if any, might have been produced in our current results had we used the [Duff and colleagues \(2007\)](#) figure scoring criteria, readers should be aware that in order to generate factor scores using the bifactor model reported herein, the scoring criteria in the RBANS Manual ([Randolph, 1998](#)) should be used. Finally, although a direct comparison of the nine competing models provided reasonably strong evidence to support the [Duff and colleagues \(2006\)](#) model, modifications were required to rectify problems (e.g., negative factor loadings) with that model. The use of these modifications could have led to results that were sample specific and which may not generalize beyond our sample.

Many of the alternative models that we compared were derived following the elimination of one or more RBANS test scores in order to meet the assumptions of factor analytic methods or because of low correlations with other items. However, it may be the case that complex models do not fit the data well because the RBANS is too brief. One alternative to reducing the number of RBANS items and factors is to explore the factor structure of the RBANS within the context of a larger test battery that provides additional variables to measure the constructs of interest. This should be a focus of future studies. Future research should also address the limitations described herein and explore the validity and clinical utility of the bifactor modification of the [Duff and colleagues \(2006\)](#) model, especially in longitudinal applications.

In summary, the results of this study contribute to the literature in five important ways. First, these results support the conclusions of [King and colleagues \(2012\)](#) and demonstrate that a model containing a global factor along with one or more specific sub-factors provides a better fit to the RBANS data than a model without a general factor. Specifically, models with a general factor were always superior to models without a general factor, and bifactor models were always superior to second-order models. Second, these results provided some support for a previously identified model for the RBANS ([Duff et al., 2006](#)). Although numerous studies have been conducted to examine the RBANS factor structure, this is the first known study to provide converging evidence for an existing model. Third, we identified slight modifications that can be made to the [Duff and colleagues \(2006\)](#) model to improve its fit; these modifications mostly retain the original structure of the [Duff and colleagues \(2006\)](#) model, but they also account for method effects and add a general factor, both of which serve to improve model fit. Fourth, our results suggest that the RBANS is best suited for measuring global cognitive abilities and visual processing. Caution is warranted when interpreting RBANS test scores, because conclusions about abilities any more specific than these two factors may lack validity. Fifth, our results can be useful to clinicians wishing to interpret RBANS scores according to the best fitting model reported here. We have developed a web-based scoring program that generates RBANS factor scores for global cognition and visual processing based on the bifactor modification the [Duff and colleagues \(2006\)](#) model. At this time, the web-based scoring is most appropriate for research, not clinical, purposes. The factor scores estimated from this model are at the interval—rather than ordinal—scale of measurement and are linearly related to the underlying abilities estimated. These characteristics are well suited for a test such as the RBANS, which was designed for making repeated measurements of cognitive functioning over time.

## Acknowledgment

Research reported in this publication was supported by the National Institute on Aging (NIA) and National Institute on Minority Health and Health Disparities (NIMHD) of the National Institutes of Health under Award Numbers R01AG039389, P30AG12300, and L60MD001849. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research was also funded in part by grants from the Hogg Foundation for Mental Health (JRG-040 & JRG-149), the Environmental Protection Agency (RD834794), and the National Academy of Neuropsychology. Project FRONTIER is supported by Texas Tech University Health Sciences Center F. Marie Hall Institute for Rural & Community Health and the Garrison Institute on Aging. The researchers would also like to thank the Project FRONTIER participants and research staff.



## References

- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Carrlozzi, N., Horner, M., Yang, C., & Tilley, B. (2008). Factor analysis of the repeatable battery for the assessment of neuropsychological status. *Applied Neuropsychology*, *15*, 274–279.
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*, 189–225.
- Cheng, Y., Wu, W. Y., Wang, J., Feng, W., Wu, X., & Li, C. B. (2011). Reliability and validity of the repeatable battery for the assessment of neuropsychological status in community-dwelling elderly. *Archives of Medical Science*, *7*, 850–857.
- 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.
- Duff, K., Langbehn, D. R., Schoenberg, M. R., Moser, D. J., Baade, L. E., Mold, J. W., et al. (2006). Examining the repeatable battery for the assessment of neuropsychological status: Factor analytic studies in an elderly sample. *American Journal of Geriatric Psychiatry*, *14* (11), 976–979.
- Duff, K., Langbehn, D. R., Schoenberg, M. R., Moser, D. J., Baade, L. E., & Mold, J. W., et al. (2009). Normative data on and psychometric properties of verbal and visual indexes of the RBANS in older adults. *The Clinical Neuropsychologist*, *23*, 39–50.
- Duff, K., Leber, W. R., Patton, D. E., Schoenberg, M. R., Mold, J. W., Scott, J. G., et al. (2007). Modified scoring criteria for the RBANS figures. *Applied Neuropsychology*, *14*, 73–83.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). 'Mini Mental State.' A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.
- Garcia, C., Leahy, B., Corradi, K., & Forchetti, C. (2008). Component structure of the repeatable battery for the assessment of neuropsychological status in dementia. *Archives of Clinical Neuropsychology*, *23*, 63–72.
- Gavett, B. E., Crane, P. K., & Dams-O'Connor, K. (2013). Bi-factor analyses of the Brief Test of Adult Cognition by Telephone. *NeuroRehabilitation*, *32*, 253–265.
- Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., & Bhaumik, D. K., et al. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, *31*, 4–19.
- Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*, 423–436.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
- King, L. C., Bailie, J. M., Kinney, D. I., & Nitch, S. R. (2012). Is the repeatable battery for the assessment of neuropsychological status factor structure appropriate for inpatient psychiatry? An exploratory and higher-order analysis. *Archives of Clinical Neuropsychology*, *7*, 756–765.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10.
- McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Technical manual. Woodcock-Johnson III normative update*. Rolling Meadows, IL: Riverside Publishing.
- Menon, C., Hall, J., Hobson, V., Johnson, L., & O'Bryant, S. E. (2012). Normative performance on the executive clock drawing test in a multi-ethnic bilingual cohort: A Project FRONTIER study. *International Journal of Geriatric Psychiatry*, *27*, 959–966.
- Mungas, D., & Reed, B. R. (2000). Application of item response theory for development of a global functioning measure of dementia with linear measurement properties. *Statistics in Medicine*, *19* (11–12), 1631–1644.
- Muthén, L. K., & Muthén, B. O. (1998–2010). Mplus user's guide sixth edition. Los Angeles, CA: Muthen & Muthen.
- O'Bryant, S. E., Hall, J. R., Cukrowicz, K. C., Edwards, M., Johnson, L. A., & Lefforge, D., et al. (2010a). The differential impact of depressive symptom clusters on cognition in a rural multi-ethnic cohort: A Project FRONTIER study. *International Journal of Geriatric Psychiatry*, *26*, 199–205.
- O'Bryant, S. E., Lacritz, L. H., Hall, J., Waring, S. C., Chan, W., & Khodr, Z. G., et al. (2010b). Validation of the new interpretive guidelines for the Clinical Dementia Rating Scale sum of boxes score in the National Alzheimer's Coordinating Center database. *Archives of Neurology*, *67*, 746–749.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879–903.
- Randolph, C. (1998). *RBANS manual*. San Antonio, TX: The Psychological Corporation.
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield unequivocal scale scores. *Journal of Personality Assessment*, *92*, 544–559.
- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16* (Suppl. 1), 19–31.
- Reitan, R., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical applications*. Tucson, AZ: Neuropsychology Press.
- Royall, D. R., Cordes, J. A., & Polk, M. (1995). CLOX: An executive clock drawing task. *Journal of Neurology, Neurosurgery, & Psychiatry*, *64*, 588–594.
- Schmitt, A., Livingston, R., Smernoff, E., Reese, E., Hafer, D., & Harris, J. (2010). Factor analysis of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) in a large sample of patients suspected of dementia. *Applied Neuropsychology*, *17*, 8–17.
- Stern, R. A., & White, T. (2003). *Neuropsychological assessment battery*. Lutz, FL: Psychological Assessment Resources.
- Wechsler, D. (2008). *Wechsler adult intelligence scale-fourth edition (WAIS-IV)*. San Antonio, TX: Pearson Assessment.
- Wechsler, D. (2009). *Wechsler memory scale-fourth edition (WMS-IV) technical and interpretive manual*. San Antonio, TX: Pearson Assessment.
- Wilde, M. (2006). The validity of the Repeatable Battery of Neuropsychological Status in acute stroke. *The Clinical Neuropsychologist*, *20*, 702–715.
- Yang, C., Garrett-Mayer, E., Schneider, J., Gollomp, S., & Tilley, B. (2009). Repeatable Battery for Assessment of Neuropsychological Status in early Parkinson's disease. *Movement Disorders*, *24*, 14533–11460.