Invariant Two Component Structure of the RBANS

Elisabeth M. Vogt
Marquette University

INVARIANT TWO COMPONENT STRUCTURE OF THE REPEATABLE BATTERY FOR THE ASSESSMENT OF NEUROPSYCHOLOGICAL STATUS (RBANS)

by

Elisabeth M. Vogt

A Thesis submitted to the Faculty of the Graduate School,

Marquette University,

in Partial Fulfillment of the Requirements for

the Degree of Master of Science in Clinical Psychology.

Milwaukee, Wisconsin

May 2015

ABSTRACT

INVARIANT TWO COMPONENT STRUCTURE OF THE REPEATABLE BATTERY FOR THE ASSESSMENT OF NEUROPSYCHOLOGICAL STATUS (RBANS)

Elisabeth M. Vogt

Marquette University, May 2015

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS: Randolph, 1998, 2012) is a brief neurocognitive instrument used to evaluate cognitive functioning in clinical settings. While this test is used regularly, investigation of the factor structure has resulted in inconsistent findings across samples. It was hypothesized that inconsistent RBANS dimensional structures are the result of methodological differences and not solely due to unique sample characteristics. The present study utilized empirically supported extraction criteria (Parallel Analysis; Minimum Average Partial Procedure) and uniformly investigated five samples. RBANS data from four samples were previously published (Carlozzi, Horner, Yang, & Tilley, 2008; Duff, Hobson, Beglinger, O'Bryant, 2010; Duff et al., 2006; Wilde, 2006) and a new clinical sample was obtained from the Gundersen Health System, Memory Center. The congruence of factor structures was investigated by conducting orthogonal vector matrix comparisons (Barrett, 2005), and a robust two factor structure reliably emerged across samples. The invariant RBANS two factor structure primarily emphasized memory and visuospatial functioning. This finding definitively clarifies the RBANS factor structure and the relationships between subtests and indices. Due to the expansive use of the RBANS, this psychometric knowledge has significant clinical implications.

ACKNOWLEDGMENTS

Elisabeth M. Vogt

Gregory Prichett, PsyD, the Gundersen Health System Memory Clinic, and the Gundersen Lutheran Institutional Review Board are thanked for their support of the present study. Thank you to Kevin Duff, PhD for providing correlation matrices for statistical analysis. Finally, thank you to James B. Hoelzle PhD for guidance and support during this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Neuropsychological assessment has a rich history that evolved from the convergence of multiple fields and continues to progress. Individuals within philosophy, science, medicine, education, art and many other disciplines have considered the relationship between brain, body and behavior in historical texts (Lezak, Howieson, Bigler, & Tranel, 2012). Today as an applied science, clinical neuropsychology focuses on the behavioral manifestation of cognitive impairment. Assessment comprises a core component of clinical neuropsychology practice. As evidence, a survey reported that 80% of neuropsychologists engage in clinical assessment at least four hours weekly and 33% spend 20 or more hours evaluating patients per week (Camara, Nathan, & Puente, 2000). Neuropsychological evaluations inform clinicians and patients of a wide variety of important diagnostic and treatment-related issues (Schoenberg & Scott, 2011).

With refinement of cognitive theories, establishment of the field of neuropsychology has increased and subsequently the standardized instruments used by clinical neuropsychologists to infer cognitive functioning evolved and increased in sensitivity (Lezak et al., 2012). In a typical, comprehensive, neuropsychological assessment multiple domains are evaluated which may include intelligence, attention, executive functioning, verbal and visual fluency, immediate memory, working memory, delayed memory, language, visuospatial ability, sensory and motor abilities, personality features, and emotional symptoms (Strauss, Sherman, & Spreen, 2006). Hundreds of standardized measures exist to evaluate many of the fore-mentioned domains. For example, the Neuropsychological Assessment Battery (NAB; Stern & White, 2003) includes multiple tasks and evaluates attention, language, memory, spatial and executive functioning in the span of four hours. Alternatively, a neuropsychological test may evaluate one specific function, such as confrontation object naming, by utilizing a measure such as the Boston Naming Test in 10 to 20 minutes (Kaplan, Goodglass, & Weintraub, 2001). Neuropsychologists reported on average that assessments typically require five hours to complete, however, this approach may not be possible or practical for many clinical populations (National Academy of Neuropsychology Board of Directors, 2007). In response to this, abbreviated testing batteries with adequate psychometric properties have been developed that are advantageous to clinicians.

The Repeatable Battery for the Assessment of Neuropsychological Status

This project aims to evaluate specific psychometric properties of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998; RBANS Update; Randolph, 2012). Development of the RBANS addressed a need for brief assessment measures that are sensitive to cognitive impairment in multiple cognitive domains. Individually administered and typically taking less than 30 minutes, it evaluates a range of cognitive abilities and has shown utility in a variety of clinical settings (e.g. see Aupperle, Beatty, Shelton, & Gontkovsky, 2002; Beatty, Ryder, Gontkovsky, Scott, McSwan, & Bharucha, 2003; Larson, Kirschner, Bode, Heinemann, & Goodman, 2005; McKay, Casey, Wertheimer, & Fichtenberg, 2007; Wilk, Gold, Humber, Dickerson, Fenton, & Buchanan, 2004).

Consideration of cognitive theory and neuropsychological functioning guided selection of specific subtests included in the RBANS (Schoenberg & Scott, 2011). These subtests are conceptually similar to popular and validated neuropsychological assessment measures and combine to create summary scores that reflect typical neuropsychological constructs (Lezak et al., 2012; Randolph, 1998). The 12 RBANS subtests contribute to five cognitive index scores (for more complete descriptions see Table 1). The Immediate Memory Index includes List Learning and Story Memory subtests, which are designed to assess auditory short-term memory and learning. A Visuospatial/Constructional Index consists of Figure Copy, to assess constructional organization, and Line Orientation, to evaluate visuospatial organization. Picture Naming, a confrontation naming task, and speed of verbal fluency, assessed with the Semantic Fluency subtest, comprise the Language Index. An Attention Index includes a simple attention task, Digit Span, and the Coding subtest, which evaluates processing speed and simple attention. The Delayed Memory Index was designed to assess temporal memory, and requires the examinee to recall previously presented stimuli presented earlier during the RBANS (i.e., List Recall, List Recognition, Story Recall and Figure Recall). An overall Total Scale index score is derived by combining all indices.

As previously mentioned, RBANS subtests parallel frequently utilized and well-validated neuropsychological measures (Camara et al., 2000). Meaningfully distinctive from corresponding traditional neuropsychological, RBANS subtests include fewer items resulting in quicker administration. For example, the RBANS Line Orientation subtest was modeled after the Judgment of Line Orientation Test (JLO; Benton, Hamsher, Varney, & Spreen, 1983). The Benton JLO test contains 30 items with only a small portion of the stimuli line drawn and takes approximately 20 minutes to complete, whereas, the RBANS Judgment of Line Orientation subtest includes 10 items with a full stimuli line and takes roughly two minutes to complete.

Table 1

Description of the Repeatable Battery for the Assessment of Neuropsychological Status

Index	Subtest	Description
Immediate Memory	List Learning	The examinee immediately recalls as many words as possible from a list of 10 semantically unrelated words presented orally repeated over 4 learning trials
	Story Memory	The examinee immediately recalls a short orally presented story over two trials.
Visuospatial/ Constructional	Figure Copy	The examinee draws a multipart geometric design while it remains displayed.
	Line Orientation	The examinee sees 13 numbered lines radiating from a single point in a semicircular fan-shaped pattern. Below that are two lines and the examinee determines what lines they match by placement and direction over ten trials with varying line sets within a time limit.
Language	Picture Naming	The examinee names 10 line drawings of common objects.
	Semantic Fluency	The examinee verbally generates as many exemplars as possible from semantic categories in 60 seconds.
Attention	Digit Span	The examiner is orally presented increasingly long strings of digits and then asked to repeat the digits in order.
	Coding	The examinee views a key with geometric shapes and corresponding numbers and fills in empty boxes below the shapes with the correct numbers in a timed task.
Delayed Memory	List Recall	The examinee recalls as many words as possible from the list presented during List Learning.
	List Recognition	The examinee hears 20 words (10 targets & 10 distracters) and asked to indicate whether each word was presented during List Learning.
	Story Memory	The examinee retells the story presented during Story Memory.
	Figure Recall	The examinee draws the figure initially copied.
Total Scale		Sum of all five indices

Source: Adapted from Groth-Marnat (2009); Randolph (1998); Strauss, Sherman, & Spreen (2012)

RBANS: Psychometric properties.

Standardized assessment measures with strong validity and reliability allow clinicians to make more accurate judgments regarding functioning. In other words, the psychometric properties of a test directly relate to its usefulness (Lezak et al., 2012). As an example, an unreliable memory test will exhibit varying degrees of association with a criterion and subsequently demonstrate little clinical or research value. A long standing area of research within the broad field of assessment pertains to the evaluation of psychometric properties of tests.

Since publication of the RBANS in 1998, multiple studies evaluated the reliability, validity, and clinical utility of the measure. In fact, a recently conducted cursory literature search identified over 1,200 studies that utilized the RBANS. The need for a clear understanding of the meaning attached to a RBANS score is further highlighted by the frequent usage of the RBANS in multiple settings. The RBANS has proved particularly useful during inpatient neuropsychological evaluations when comprehensive testing is not practical (Lezak et al., 2012). While this measure was originally developed for dementia evaluations, clinicians have utilized the RBANS as a key aspect of assessment across multiple clinical populations such as those presenting with Parkinson’s disease (Beatty et al., 2003), stroke (Larson et al., 2005), multiple sclerosis (Aupperle et al, 2002; Beatty, 2004), schizophrenia (Wilk et al., 2004) and traumatic brain injury (McKay et al., 2007), among others. Consistent with literature investigating traditional neuropsychological tests, individuals with clinical conditions invariably perform worse on the RBANS subtests than the RBANS normative sample.

Indicative of the integration of this measure into neuropsychological practice, the RBANS served as a “gold standard” in a research study that evaluated the negative predictive power and positive predictive power of novel, brief, computerized neuropsychological assessment (Woodhouse et al., 2013).

Reliability

An important psychometric property, reliability impacts the utility of a measure. In general, reliability reveals the consistency of measurement (Slick, 2006). Defined several different ways, reliability statistics include: internal consistency, consistency over time, consistency across alternate forms, and consistency across raters. Reliability provides some indication of the error (the degree of and sources of variability that influence a test score) associated with a specific test score (Slick, 2006). Traditional benchmarks for reliability coefficients are suggested as follows: very high +.90, high .80 to .89, adequate .70 to .79, marginal .60 to .69, and low < .59 (Slick, 2006). Types of reliability are explained in the following paragraphs with related RBANS empirical findings.

Internal reliability (also known as internal consistency) conveys the degree to which different items of the same measure are correlated. It is typically conveyed by reporting split-half reliability coefficients or coefficient alpha. Split-half reliability is established by dividing a test in two and evaluating the association between the two halves. On average, across age groups RBANS internal consistency, determined through split-half reliability (Spearman-Brown), was reported to be .80 (Randolph, 1998). Reliability coefficients of the Total Scale were high (.86 to .94), but the individual indices were significantly lower (range .55 to .78; Hobart, Goldberg, Bartko, & Gold, 1999; Randolph, 1998).

Coefficient alpha indicates the degree to which a set of items measures a single dimension (as opposed to the association between parts of a test). McKay and colleagues (2007) investigated the internal consistency of RBANS indices in a sample of patients who had sustained traumatic brain injuries and reported a wide range of alpha coefficients. While the Total Score (α = .83), Delayed Memory (α = .77), Visuospatial/Constructional (α = .76), and Immediate Memory (α = .75) indices exhibited good internal consistency, the remaining RBANS indices had unacceptable internal consistency (Attention α = .16; Language α = .33). Ultimately, this raises the question of whether select indices (e.g., Attention and Language) evaluate a single latent construct.

Test-retest reliability describes the stability of measurement when the same test is administered to a single individual at different points in time. A test with good temporal stability minimally changes for normal individuals that are not experiencing cognitive decline. With respect to the RBANS, Duff and colleagues (2005) investigated the stability of RBANS index and subtest scores over a period of one year. Utilizing a sample of 455 “typically aging” adults over 65 years, it was reported that the Total Score was most stable (.83) and individual indices varied significantly. Test-retest reliability of indices ranged from low (Language = .53) to adequate (Total Score = .83). Evaluation of test-retest reliability of subtests demonstrated similar variability and ranged from low (Figure Copy = .51) to adequate (Coding = .81).

A novel feature of the RBANS, relative to many other neuropsychological measures and batteries, is that alternate forms have been published for serial evaluation.

The advantage of alternate forms is that a repeat assessment could be conducted while minimizing (but not eliminating) the confounding variable of practice effects (Randolph, 1998). Alternate form consistency, between Form A (the form most frequently administered by clinicians) and Form B of the RBANS, was reported by Randolph (1998) to be high for the Total Score (.82), but again variable for indices (ranging from Language r =.46 to Attention r = .80). Two follow-up studies with patients who had schizophrenia revealed a similar alternate form reliability pattern with the Total Score demonstrating excellent reliability (r = .84) and other indices varying widely. The Language Index demonstrated the lowest stability, whereas the Attention Index demonstrated the highest reliability (Wilk et al., 2002: Language r =.36 and Attention r =.76; Gold et al., 1999: Language r = .56, and Attention r = .91). Overall, given the varying alternate form reliability coefficients across RBANS indices, it is recommended that *only* the Total Scale index score be utilized to evaluate change in cognitive functioning over time (Groth-Marnat, 2009; Strauss et al., 2006).

Interrater reliability is also important to consider because it explains the amount of variance in scores due to examiner judgment, or in other words, this reliability evaluates the consistency of administration and scoring (Slick, 2006). Evaluation of the interrater reliability of the Design Copy and Design Memory subtests was investigated because those subtests include somewhat subjective scoring criteria. Randolph's (1998) report of inter-rater consistency of the Figure Copy and Figure Recall were acceptable and reported as identical reliability coefficients (r = .85). An alternative scoring system has been established for these subtests in response to researchers' concerns that individuals were obtaining scores lower than expected (Duff, Patton, Schoenberg, Mold, Scott, & Adams, 2003; Gontkovsky, Beatty, & Mold, 2004). The interscorer reliability of the modified criteria is higher than the original criteria developed by Randolph (Figure Copy $r = .94$; Figure Recall $r = .98$; Duff, Leber, Patton, Schoenberg, Mold, Scott, & Adams, 2007).

Reliability is important to consider when selecting tests because it impacts the standard error of measurement (SEM). SEM indicates the amount of error that is associated with measurement, and determines the degree to which a specific score might fluctuate for a single person (Slick, 2006). The SEM of a score is inversely related to the reliability of the measure, so as reliability increases SEM decreases. RBANS index scores SEM values varied, ranging from 3.84 to 6.65 (Randolph, 1998). By definition, those Index scores with poorer reliability (Visuospatial/Construction and Language) had the largest SEM values (6.65 and 6.52, respectively). The overall composite score exhibits the strongest reliability (Total Scale SEM = 3.84) supporting previously mentioned reports that this index is most stable at single evaluation points and in assessing cognitive change over time.

Validity

The fore-mentioned types of reliability (e.g. consistency of the RBANS) provide necessary framework to evaluate the validity of the RBANS (e.g. accuracy of construct assessment). Validity provides the property of meaning attached to a test score (Slick, 2006). The concept of validity is often incorrectly described simply as whether or not a test measures what it is intended to measure. More specifically, validity refers to the appropriateness or accuracy of the interpretation of test scores (Slick, 2006). There are certainly situations when a valid measure will not be appropriate to use in a specific context (e.g., using an intelligence test validated with English speaking adults with a Spanish speaking student). There is a relationship between validity and reliability: a valid measure must be reliable, but the inverse is not true.

Validation of a test is a continual process, and it is believed that validation of measures is not only the responsibility of the test developer but also those that utilize the test in clinical practice and research (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Messick (1995) proposed a comprehensive model of validity in which six separate, distinguishable types of evidence contribute to validity in order to create evidence for interpretation of a measure (content related, substantive, structural, generalizability, external, and consequential evidence sources). The Standards for Educational and Psychological Testing (1999) propose a similar model that includes: evidence based on test content, response processes, internal structure, relations to other variables, and consequences of testing. However, the inclusion of consequences of testing as evidence for validity is frequently criticized as too far reaching (Slick, 2006). While many models of validity exist, the most commonly seen is a tripartite model that includes: content validity, criterion-related validity, and construct validity (Slick, 2006). The tripartite model of evidence for validity and related RBANS empirical literature will be discussed.

Content validity refers to the quality of test measure in relation to the relevance representativeness of the test content. The RBANS content was based on a theoretical model of cognition and supported by use of tasks that are similar to other well validated measures. For example the RBANS includes a verbal fluency task (Semantic Fluency) that is similar to the Controlled Oral Word Association test (Benton, Hamsher, & Sivan, 1994), a visual perception task (Line Orientation) which is similar to the JLO test, and other tasks that are shortened versions of empirically validated measures. Utility of abbreviated versions of these longstanding measures demonstrated to assess specific cognitive constructs (e.g., verbal fluency) clearly suggests content validity (Randolph, 1998).

Criterion-related validity encompasses concurrent and predictive validity. Concurrent validity is important for neuropsychological test measures used to identify cognitive impairment associated with specific disorders. In other words, concurrent validity demonstrates the clinical sensitivity of the measure. Predictive validity refers to the ability of the measure to accurately inform a clinician of possible future outcomes (Slick, 2006). At the time of development, the clinical sensitivity and clinical utility of the RBANS were investigated with adults that had various neurological and psychiatric disorders (Randolph, 1998). In brief, it was reported that Index Score patterns varied as expected based upon cognitive profiles typically associated with differing neurocognitive impairment in clinical samples of individuals with Alzheimer’s disease, Vascular Dementia, Mixed Dementia, Huntington’s Disease, Parkinson’s Disease, Depression, Schizophrenia, or Traumatic Brain Injury.

In addition, the ability of RBANS scores to accurately predict return to work, instrumental activities of daily living, and disability outcomes has been investigated. Predictive validity of the RBANS was demonstrated when clinical outcomes of patients that experienced a stroke were accurately predicted at 12 months status-post stroke

(Larson et al., 2005). Specifically, Larson and colleagues (2005) determined that the RBANS Total Score, Language, Immediate Memory, Delayed Memory, and Visuospatial/Construction indices demonstrated predictive validity in stroke patients due to strong, positive correlations with cognitive disability after one year. Notably, the Attention Index was not correlated with disability outcome.

In recent years, numerous researchers have additionally provided empirical evidence for concurrent validity of the RBANS. Specifically, Index scores were found to demonstrate distinct and reliable patterns in normal and psychiatric samples demonstrating the clinical utility of the RBANS to distinguish impairment from non-impairment (Gold et al., 1999; Hobart et al., 1999; Iverson, Brooks, & Haley, 2009; Wilk et al., 2002). Researchers have also further demonstrated the clinical utility of this measure with various neurological and psychiatric disorders such as Alzheimer's disease (Randolph, Tierney, Mohr, and Chase, 1998), Parkinson's Disease (Beatty et al., 2003), stroke (Larson et al., 2003), and general cognitive decline (Duff et al., 2008) in which the RBANS displayed the pattern of performance expected for each clinical population. Each of these specific clinical populations demonstrated distinct patterns of results on the RBANS, demonstrating the range of domains measured and clinical efficacy of the measure.

Construct validity of a test measure is determined through multiple ways, including: evaluation of convergent and divergent validation, and component/factor identification through factor analysis. Overall construct validity of the RBANS was originally demonstrated with convergent and discriminant validity of the RBANS in correlational analyses with other commonly used neuropsychological assessments

(Randolph, 1998). RBANS indices converged with measures of intelligence, memory, language, attention, and executive functioning in an expected manner (Gold, Queern, Iannone, & Buchanan, 1999; Hobart et al., 1999; Larson et al., 2005; McKay et al., 2007; Pachet, 2007; Randolph, 1998).

Factor Structure of the RBANS

The internal or underlying structure of the RBANS has been investigated by researchers who have sought to evaluate RBANS construct validity. A primary goal in determining the factor structure of a neuropsychological assessment measure is to summarize relationships between variables (e.g. RBANS subtests) in order to define the underlying dimensions, which are then inferred to reflect cognitive constructs (Tabachnick & Fidell, 2013). The factor structure is important because it informs the fidelity of the scoring structure to the construct assessed by the test (Messick, 1995). A clearly defined factor structure helps clinicians evaluate the construct validity of the test and directly affects the credibility of the measure in clinical decision making (King, Bailie, Kinney, & Nitch, 2012).

When the RBANS was published, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were not reported in the manual (Randolph, 1998). To date, six studies investigating the factor structure of the RBANS have provided inconsistent results, which has left clinicians to question what constructs are evaluated and the validity of the Index structure (Carlozzi, Horner, Yang, and Tilley, 2008; Duff et al., 2006; Garcia, Leahy, Corradi, & Forchetti, 2008; King et al., 2012; Schmitt, Livingston, Smernoff, Reese, Hafer, & Harris, 2010; Wilde, 2006). In the following sections, factor analysis and related methodological decisions will be elaborated upon to explain convergent and divergent results that have appeared in the literature.

Confirmatory factor analysis

In CFA, theory dictates what factor structure should be observed. The “fit” between a hypothesized factor structure and the actual data is then evaluated (Tabachnick & Fidell, 2013). In other words, researchers specify how specific items (e.g., subtests) relate to assumed theoretical constructs. Three CFA studies have been conducted to investigate the RBANS factor structure (Carlozzi et al., 2008; Duff et al., 2006; King et al., 2012). Each study has evaluated whether the underlying factor structure of the RBANS was consistent with the RBANS Index structure. Both a 5 factor structure to mirror the index organization and a single factor structure to replicate the overall score were investigated (Carlozzi et al., 2008; Duff et al., 2006; King et al., 2012). Across diverse samples, including community dwelling older adults (Duff et al., 2006), veterans referred to a memory disorder clinic (Carlozzi et al., 2008), and patients with psychiatric disorders (King et al., 2012), CFA results have not supported a five or one factor structure. Notably across studies, immediate and delayed memory indices were highly correlated, which contributed to a misfit between the underlying structure and expectation. This is not surprising given that numerous factor analytic studies investigating memory have found a single memory dimension that encompasses both immediate and delayed memory (Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003; Dowling, Hermann, La Rue, & Sager, 2010; Hoelzle, Nelson, & Smith, 2011).

It is noteworthy that some researchers have expressed concern that CFA might not be an ideal method to evaluate the construct validity of measures (Lee & Aston, 2007). It has been observed that traditional fit indices (e.g., χ^2 test) reject models that are only trivially misspecified when the sample size is large (Bentler & Bonett, 1980). Additionally, confirmatory factor analysis may lack sensitivity for relationships between variables that may be highly discreet or complex since these must be specified by the researcher a-priori (Hoelzle & Meyer, 2013). As an illustration, this is a possible explanation for why omnibus Big Five personality inventories have not replicated when evaluated with CFA models (Church & Burke, 1994; Gignac Bates, & Jang, 2007; McCrae et al., 1996), despite the influence of factor analytic methods on the development of the Big Five model of personality. Previously described RBANS CFA studies should be interpreted with this in mind. In other words, the failure of CFA methods to support specified models does not necessarily mean the battery is invalid: rather, it raises questions about the relationship between subtests and composition of indices. This conclusion suggests alternative methods should be considered to evaluate construct validity.

Exploratory factor analysis

EFA is an alternative method to evaluate construct validity. In contrast to CFA, which is theory driven, EFA is a data driven method of variable reduction where multiple variables (e.g. subtests) are organized into factors or components that reflect relationships (e.g. cognitive constructs) between the variables (Goldberg & Velicer, 2006). EFA methods have been utilized to investigate the RBANS factor structure 6 times (see Table

2; Carlozzi et al., 2008; Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010; Wilde, 2006). A cursory review of this broad literature provides evidence that different factor structures have been reported, which clearly raises the question of whether the RBANS has an invariant factor structure.

Table 2

RBANS Factor Analytic Studies Overview

Study	Sample	Subtests Analyzed	Method	Rotation	Extraction Criteria	Latent Constructs (% Variance Explained)
Wilde (2006)	210 Patients with CVA	12	PCA	Varimax	EV > 1 Scree Plot	1. Language/ Verbal Memory (37%) 2. Visual / Visual Memory (24%)
Duff et al. (2006)	824 Normal Aging Adults	9	CFA ML EFA	Varimax Promax	EV > 1 Scree Plot	1. Verbal Memory 2. Visual Processing (60% Combined)
Garcia et al. (2008)	351 Memory Clinic Patients	12	PCA	Direct Oblimin	EV > 1 Scree Plot	1. Memory (39.5%) 2. Visuomotor Processing (13.51%) 3. Verbal Processing (8.42%)
Carlozzi et al. (2008)	175 Memory Clinic Patients	11	CFA ML EFA	Varimax	Chi-Square Test Variance Explained	1. Memory, visual motor, verbal fluency (89.4%) 2. Visuospatial & Attention (10.6%)
Schmitt et al. (2010)	636 Memory Clinic Patients	12	PCA PAFA	Varimax Promax	EV > 1 Scree Plot	1. Memory & Learning 2. Visuospatial & Attention (54.4% Combined)
King et al. (2012)	167 Patients with SCZ	12	CFA PAFA PCA	Promax	EV > 1 Scree Plot SE of Scree Horn's PA MAP	1. Memory (13.9%) 2. Speed of Processing (8.2%)

Note. CVA = Cerebral vascular accident; SCZ = Schizophrenia; PCA = Principal components analysis; ML = Maximum likelihood; EFA = Exploratory factor analysis; EV = Eigenvalue; CFA = Confirmatory factor analysis; PAFA = Principal axis factor analysis; SE of Scree = Standard error of the scree plot; PA = Parallel analysis; MAP = Minimum average partial

Inspection of the pattern of factor loadings in Table 3, reveals similarities and discrepancies across studies. Published factor loadings with methodological similarities are grouped accordingly in Table 3. Importantly, actual values of factor loadings vary dependent upon methodology utilized (e.g. PCA vs. ML EFA, rotation) so specific loadings cannot be equated across all samples (Tabachnick & Fidell, 2013). Nevertheless, across studies, it appears that the key primary loadings on one factor typically reflect memory functioning (List Learning, Story Memory, List Recall, List Recognition, Story Recall). The Figure Recall subtest loading varies across studies between a first primarily memory factor and second factor typically reflecting visuospatial abilities or attention. The greatest discrepancy across studies is how processing speed, language and attention tasks are associated with factors.

Table 3

RBANS Factor Analytic Study Factor Loadings and Eigenvalues

RBANS Subtests	Wilde (2006)[a]		Schmitt et al. (2010)[a]		King et al. (2012)[a]		Duff et al. (2006)[b]		Carlozzi et al. (2008)[b]		Garcia et al. (2008)[c]		
	1	2	1	2	1	2	1	2	1	2	1	2	3
List Learning	**.84**	.14	**.85**	-.05	**.85**	-.05	**.66**	.27	**.66**	**.42**	.24	-.13	**.54**
Story Memory	**.75**	.07	**.65**	.16	**.65**	.16	**.76**	.23	**.85**	.26	**.56**	.11	.36
Figure Copy	.02	**.92**	.02	**.53**	.02	**.53**	.11	**.64**	.26	**.65**	.04	**.91**	-.14
Line Orientation	.11	**.82**	.01	**.59**	.01	**.59**	.14	**.53**	.23	**.82**	-.05	**.85**	.05
Picture Naming	**.67**	.05	-.07	**.57**	-.07	**.57**	–	–	.37	**.52**	.19	.08	**.55**
Semantic Fluency	**.70**	.21	-.01	**.53**	-.01	**.53**	–	–	**.55**	**.40**	.20	.29	**.51**
Digit Span	**.48**	.07	.16	.30	.16	.30	–	–	.30	.34	-.22	.08	**.71**
Coding	.41	**.71**	.21	**.46**	.21	**.46**	.31	**.53**	**.59**	**.62**	.11	**.67**	.27
List Recall	**.74**	.22	**.86**	-.11	**.86**	-.11	**.71**	.20	**.67**	.25	**.84**	-.08	.01
List Recognition	**.78**	.15	**.74**	-.02	**.74**	-.02	**.56**	.19	**.59**	**.42**	**.59**	.05	.23
Story Recall	**.77**	.24	**.71**	.20	**.71**	.20	**.80**	.25	**.77**	.33	**.87**	.01	.01
Figure Recall	.23	**.79**	.28	**.44**	.28	**.44**	.37	**.57**	**.56**	**.49**	**.82**	.16	-.20
Eigenvalue	5.33	1.98	5.39	1.06	5.39	1.06	4.09	1.29	17.2	2.04	4.74	1.62	1.01

Note. Primary factor loadings are in boldface. Duff (2006) excluded Digit Span, Picture Naming, and Semantic Fluency subtests from analyses.
[a] = Pattern matrix factor loadings after varimax rotation in PCA. [b] = Factor loadings after varimax rotation in maximum likelihood EFA. [c] = Pattern matrix factor loadings after direct oblimin rotation in PCA.

Researchers have posited that the previously described inconsistent RBANS factor structures reflect sample specific differences (see Table 4; Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010; Wilde, 2006). It has been argued that underlying sample characteristics (e.g. normal cognitive functioning, memory impairment, psychiatric diagnoses) obscure the underlying cognitive constructs that may emerge in EFA and therefore impact the resulting solution (Delis, Jacobson, Bondi, & Hamilton, 2003). Garcia and colleagues (2008) reported a 3 factor EFA solution from a mixed clinical sample of outpatients with memory disorders and suggested this solution differed in terms of sample characteristics when compared to other RBANS factor analytic studies. Duff and colleagues (2006) and King and co-authors (2012) offered a highly similar explanation for factor solution differences. Simply stated, authors of previous factor analytic studies proposed that solutions vary as a function of underlying sample characteristics. However, as demonstrated in Table 2, researchers found very similar solutions with clinical and non-clinical groups suggesting that alternative factors (e.g. methodology) might contribute to subtle solution discrepancies.

Table 4

Sample Characteristics of Published RBANS Factor Analytic Studies

Study	Sample	Gender	Age (SD)	Ethnicity/Race
Wilde (2006)	210 Clinical Inpatients with CVA in Rehab Unit	50.5% Female 49.5% Male	61.91(13.97)	59.5% Caucasian 41.9% African American 7.6% Hispanic 1.0% Asian
Duff et al. (2006)	824 Non-Clinical Community Dwelling Adults	57% Female 43% Male	73.4(5.8)	86% Caucasian
Garcia et al. (2008)	351 Clinical Outpatients with Memory Disorders	58.7% Female 41.3% Male	77.9(7.5)	99% Caucasian
Carlozzi et al. (2008)	175 Clinical Outpatient Veterans in VA Memory Center	0% Female 100% Male	74.1(8.0)	71.4% Caucasian 28.6% African American
Schmitt et al. (2010)	636 Clinical Outpatients with Dementia or MCI	60.9%Female 39.1% Male	76.61(7.29)	88% Caucasian 4% African American 1% Hispanic 0.5% Asian American 7% Unknown
King et al. (2012)	167 Clinical Inpatients with Schizophrenia	11.4% Female 88.6% Male	42.76(9.73)	44% Caucasian 27% African American 14% Hispanic/Latino 6.6% Multiethnic 4.8% Asian/Pacific Islander 4.2% other

The belief that sample based differences might ultimately impact the factor structure, is consistent with ideas put forth by Delis, Jacobson, Bondi, & Hamilton (2003). They investigated the factor structure of California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1994, 2000) in samples of (a) healthy participants, (b) individuals with Alzheimer's disease (AD), and (c) individuals with Huntington's disease. The CVLT assesses immediate and delayed memory through a verbally administered word list so based upon cognitive theory the expectation was that a two factor solution reflecting immediate and delayed memory would emerge. The CVLT factor structure differed in the clinical sample of patients with Alzheimer's disease. A one factor solution was present in that sample, whereas a two factor solution was observed in the two other groups.

Based on the previously described findings, Delis and colleagues (2003) concluded that utilization of factor analysis for validity testing was an "outdated approach." In response to this position, Larrabee (2003) clarified why different solutions emerged when the CVLT factor structure was investigated across different samples. He highlighted that the sample of patients with AD had memory issues that could be characterized as rapid forgetting, which may have produced a floor effect that confounded results. Larrabee reiterated that factor analysis is an important method to evaluate clinical tests and highlighted the importance of careful subject selection and attention to methodological decisions. In regards to the current measure of focus, two samples from memory clinics demonstrated that immediate and delayed memory tasks loaded onto a primary memory component (Carlozzi et al., 2008; Garcia et al., 2008). In both studies individuals were not separated into groups based upon diagnosis (e.g. disorders with significant delayed memory impairment) which likely prevented a floor effect. Duff and colleagues (2006) demonstrated this same pattern in which immediate and delayed memory tasks loaded onto a single factor in a non-clinical sample.

The issue of whether analyzing patient and non-patient samples should result in consistent factor solutions has been thoroughly explored by researchers interested in measures that quantify mood and personality features. For example, O'Connor (2002) investigated the factor structure of 37 different personality and psychopathology measures. Multiple clinical and non-clinical samples were identified and each sample was factor analyzed using empirically supported methods (described in greater detail below). O'Connor (2002) conclusively identified that factor structures generally replicated across clinical and non-clinical samples for each measure when appropriate methods were utilized. A similar finding was reported by Hoelzle and Meyer (2009) where an invariant factor structure underlying the Personality Assessment Inventory (Morey, 1991) was reported across clinical and non-clinical samples. Therefore, while researchers purport that different samples often yield different factor structures, it appears that this variability may actually reflect methodological decisions made by researchers and not underlying sample characteristics. The following sections will briefly describe methodological issues that may be contributing to inconsistent factor solutions across different samples.

Extraction Method

It is often overlooked that there are multiple way to conduct EFA. Factor analysis (FA) and principal components analysis (PCA) are both data driven approaches to identify underlying dimensions, but they differ in theory. Traditional FA extracts *factors* that are comprised of common variance, whereas PCA extracts *components* that consist of unique, shared, and error variance. Mathematically, the primary difference is what value is placed on the main diagonal of the correlation matrix (Goldberg & Velicer, 2006; Tabachnick & Fidell, 2013). In FA, the covariance between variables is analyzed and error and unique variance is excluded: values in correlation matrix diagonal are communalities of the shared variance between variables (e.g., values between 0 and 1; Tabachnick & Fidell, 2013). In PCA, ones are in the diagonal of the correlation matrix and all variance, including error and unique variance, is disseminated to the components (Tabachnick & Fidell, 2013). Since error and unique variance are omitted in FA, the observed variables and observed correlation matrix are not fully reproduced, the factors are approximates.

Tabachnick and Fidell (2013) advise that if the research goal is to determine a theoretical solution without variability influenced by error and unique variance then FA should be selected, whereas, PCA will produce a unique mathematical solution accounting for test score error. On the other hand, others suggest the difference between the two methods does not meaningfully impact results (Goldberg & Velicer, 2006; Hoelzle & Meyer, 2013). Consistent with this position, two RBANS factor analytic studies reported that when both FA and PCA were conducted similar results were obtained (Carlozzi et al., 2008; King et al., 2012). This suggests that decisions pertaining to extraction method are unlikely to account for differences observed when reviewing RBANS factor analytic studies.

Extraction criteria

An important methodological decision when conducting factor analysis is to determine how many factors will be retained. Four previous factor analytic RBANS studies (Duff et al., 2006; Garcia et al., 2008; Schmitt et al., 2010; Wilde, 2006) utilized the two most common methods to determine factor retention, factors with eigenvalues greater than one (i.e., *Kaiser's criterion*; Kaiser, 1960) and visual examination of a scree plot (Cattell, 1966). Carlozzi and colleagues (2008) reported that criteria for judging the number of factors to extract included investigation of Maximum Likelihood and chi square test statistics, which ultimately resulted in a two factor solution.

The methods used by researchers to investigate the RBANS factor structure are somewhat inconsistent with best practice guidelines (Fava & Velicer, 1992a, 1992b; Goldberg and Velicer, 2006; Hoelzle & Meyer, 2013; Hubbard & Allen, 1987; Zwick & Velicer, 1982, 1986). In brief, empirical research suggests that multiple methods of factor extraction should be utilized in order to identify a reliable factor solution which include: interpretation of the scree plot, Horn's (1965) parallel analysis, and the Minimum Average Partial (MAP) Procedure (Velicer, 1976). King and colleagues (2012) are the only researchers that followed these recommendations for factor extraction. They utilized five methods (Kaiser's criterion, interpretation of the scree plot, parallel analysis, MAP procedure, and evaluation of the Standard Error of Scree) to determine the number of factors in the solution. These guidelines did not converge; Kaiser's criterion indicated that two factors should be retained and all other methods suggested a one factor solution. Despite converging evidence that one factor should be retained, King (2012) selected a final solution that was supported only by Kaiser's criterion, a factor retention strategy that is not supported by empirical evidence. It is a significant issue that researchers have not uniformly utilized empirically-supported guidelines to determine how many factors to retain. Research suggests that neglect of empirical guidelines for factor retention might result in inconsistent findings across studies (Hoelzle & Meyer, 2009; O'Connor, 2002). It is possible that if empirically-supported procedures were implemented, an invariant factor structure may emerge.

Rotation

Determining how extracted factors will be rotated prior to interpretation is also an important methodological decision, and recommendations clearly indicate that when factors (e.g., distinct cognitive constructs) are known to be correlated oblique rotation (e.g. Direct Oblimin) should be selected (Tabachnick & Fidell, 2013). Interestingly, only one RBANS factor analytic study utilized an oblique method of rotation (Garcia et al., 2008). Four of the prior studies (Carlozzi et al., 2008; Duff et al., 2006; Schmitt et al., 2010; Wilde, 2006) utilized orthogonal rotational (e.g. Varimax), which assumes that factors are uncorrelated. This is a questionable decision because by nature cognitive constructs are correlated with each other (e.g., attention is meaningfully related to memory functioning). Researchers likely selected varimax rotation because it often results in easily interpreted solutions by attempting to maximize high and minimize small loadings (Hoelzle & Meyer, 2013). A third rotation, Promax, which involves aspects of oblique and orthogonal rotation, was utilized in three studies (Duff et al., 2006; King et al., 2012; Schmitt et al., 2010). In short, this procedure rotates orthogonal factors to oblique positions (Tabachnick & Fidell, 2012). Promax rotation, while typically referred to as an oblique rotation actually appears to be more similar to a basic orthogonal rotation.

It is unclear how researchers' decisions to use either orthogonal or oblique rotation might impact findings. If obliquely rotated factors are not highly correlated, they will approximate an oblique solution. This likely explains why Duff and colleagues (2006) and Schmitt and colleagues (2010) reported that varimax and promax rotations resulted in similar solutions. On the other hand, to the degree that obliquely rotated factors are highly correlated, the solution is likely to diverge with an orthogonally rotated solution. In any event, there is a strong theoretical rationale for using oblique rotation given the well documented relationships between cognitive abilities.

Current Study and Significance

Based upon a review of literature, it is clear that discrepant RBANS factor structures have been reported. A common factor emerges across studies that reflects the latent construct of memory, but questions remain as to whether an invariant factor structure might be present. While many authors (Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010; Wilde, 2006) have suggested discrepant findings are related to sample-based issues, there is a body of literature that suggests methodological issues, specifically factor retention decisions, may meaningfully contribute to these differences (Hoelzle & Meyer, 2009; Larrabee, 2003; O'Connor, 2002). The overarching goal of the present study was to evaluate whether an invariant RBANS factor structure might emerge after systematically analyzing different RBANS datasets using empirically supported methods (Goldberg & Velicer, 2006; Hoelzle & Meyer, 2013; Tabachnick & Fidell, 2013). If a replicable factor structure is identified, novel construct component scores (i.e., empirically derived composite scores) could be generated that would offer clinically relevant information about patients' cognitive functioning. Theoretically, these scores should be more reliable and provide clinically relevant information regarding an individual's neurocognitive functioning. Important follow-up research might then evaluate the incremental gain of using empirically-based factor scores over traditional RBANS index scores in identifying cognitive symptoms associated with neurologic and psychiatric conditions.

To achieve this goal, the present study sought to obtain RBANS data from multiple adult samples and proposed that a consistent factor structure might emerge between several adult clinical and non-clinical samples. The congruence of factor solutions could then be investigated by conducting orthogonal vector matrix comparisons in order to determine whether a structure reliably emerges across samples (Barrett, 1986). The outcome of this study could clarify the factor structure of the RBANS, the relationships between subtests and indices, and the construct validity of this measure. Due to the expansive use of this neuropsychological instrument, a definitive conceptualization of this instrument may have significant clinical implications in that it would clarify the relationships between subtests and indices. In other words, it would foster more accurate interpretation of RBANS data.

Method

The present study sought to reanalyze previously published RBANS data and evaluate a new clinical sample that has not yet been investigated. The latter sample consists of archival clinical data obtained from a memory clinic (Gundersen Health System) and is described below. Carlozzi and colleagues (2008) published their RBANS correlation matrix so it was possible to include that data in analyses. Wilde (2006) had previously supplied the correlation matrix from his RBANS factor analytic study for a prior research project (Hoelzle, 2008) so that sample is also included in analyses. Additionally, a literature review was conducted to locate additional published RBANS subtest correlation matrices utilizing raw scores. Over 50 articles were reviewed, and no additional matrices were located.

Through personal communication, the correlation matrices describing the relationships between RBANS subtests were requested from each of the remaining four corresponding authors of the previously published RBANS factor analytic studies (Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010). Kevin Duff graciously supplied the RBANS correlation matrix that was previously analyzed (Duff et al., 2006) and numerous other correlation matrices, of which one sample was of sufficient size for further analyses (Duff, Hobson, Beglinger, O'Bryant, 2010). The remaining authors did not provide correlation matrices, so it was not possible to investigate those samples (Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010).

Samples and Procedures

Samples.

The samples are independently described in the following sections. Archival data from patients assessed in the Gundersen Health System Memory Center in La Crosse, Wisconsin was obtained and analyzed. Institutional Review Board Approval was obtained for this archival study from both Gundersen Health System and Marquette University. The author of this study collected and de-identified the neuropsychological data and entered all testing results into SPSS version 20 database (SPSS, Inc., Chicago, IL). The patients within this sample were evaluated by a multidisciplinary team in a comprehensive memory assessment clinic. The Gundersen Health System Memory

Center sample included 393 patients who were evaluated between January 1, 2009 and June 1, 2013. Participants with significant cognitive impairment [e.g., Mayo Short Test of Mental Status (Kokmen, E., Naessens, J. M., & Offord, K. P., 1987) score <14 or severe intellectual disability, *n* = 48] were administered an abbreviated neuropsychological battery that did not include the RBANS and are therefore excluded from this study. Patients included in this study (*n* = 345) ranged in age from 44 to 96 years (mean = 75.29, SD = 8.68). Fifty-three percent (*n* = 186) of this sample was female. Estimates of premorbid intellectual functioning indicated this sample was within the average range [Wechsler Test of Adult Reading (WTAR; Psychological Corporation, 2001) *n* = 130, *M* = 95.43(15.66); ACS Test of Premorbid Functioning (TOPF; Wechsler, 2009) *n* = 217, *M* = 94.08(11.23)]. The majority this sample completed high school [*M* = 12.66(3.10)]. This sample was diverse diagnostically, though the majority of patients received a diagnosis of dementia (Alzheimer's Disease 24.9%, Dementia NOS 18.8%, Cognitive Disorder NOS 13.2%, Vascular Dementia 10.9%, Frontotemporal Dementia 7.6%, Mild Cognitive Impairment 7.1%, Normal/No Impairment 5.1%, Mixed Dementia 5.1%, Lewy Body Dementia 2.3%, Parkinson's Dementia 1.5%, Pervasive Developmental Disability 1.5%, ADHD 0.8%, Wernicke-Korsakoff's 0.3%). Racial and ethnic identity was not reliably available for this sample in electronic medical records, though the sample was predominantly Caucasian and not of Hispanic origin.

A target sample size of 300 was selected based upon a review of published benchmark recommendations of sample size for PCA. Based upon empirical literature review, Hoelzle and Meyer (2013) reported that each of the following have been recommended as sufficient sizes; 100 to 150 participants (Gorsuch, 1983; Kline, 1979), 200 to 250 participants (Cattell, 1978; Guilford, 1954), 300 participants (Tabachnick & Fidell, 2012), or 500 participants (Comrey & Lee, 1992). Velicer and Fava (1998) empirically investigated the effect of various sample sizes (e.g., 50, 100, 150, 200, 400, or 800) on factor loadings, and identified that low, but non-trivial, loadings (.40) were significantly impacted by smaller sample sizes (e.g., 50-200). Previous factor analytic studies of the RBANS have demonstrated some low primary factor loadings (see Table 1). Based upon review of these recommendations, the intended enrollment for the new sample was to be at least 300 participants. Additionally, this sample size is in line with previously published RBANS factor analytic studies that have included 167 to 864 participants (see Table 4).

All remaining patients (*N* = 345) underwent a comprehensive neuropsychological evaluation at time of diagnosis. This intentionally selected clinical sample demonstrates homogeneity in some criterion (e.g., age range) and heterogeneity in other criterion (e.g., resulting diagnosis). The balance of homogeneity and heterogeneity of a sample in a factor analytic study is important for generalizability (Goldberg & Velicer, 2006). Furthermore, this specific memory center sample was selected in order to ensure that variables exhibit a spread in scores necessary for correlations to be strong and subsequent factors to emerge in the analysis (Tabachnick & Fidell, 2013). Additionally, this sample was selected in order to avoid the occurrence of floor effects (e.g., scores that cluster at the lowest values possible) or ceiling effects (e.g., majority of scores at the highest end of the distribution) since restriction in range directly affects the strength of factor loadings and strength of correlations (Fabrigar et al., 1999). For ease of identification, in subsequent writing and tables this novel clinical sample will be referred to as the "Vogt sample."

Sample characteristics of the previously published studies are briefly described in this section. Carlozzi and colleagues (2008) investigated the factor structure of the RBANS utilizing data from 175 veterans seen in a memory clinic. Patients within the Carlozzi (2008) sample were on average 74.1 (8.0) years old, primarily Caucasian (71.4%), male (100%), and had 11.3 (4.0) years of education. Diagnosis resulting from a comprehensive memory evaluation varied (Cognitive Disorder NOS 23.2%, Alzheimer's Disease 13.7%, Mild Cognitive Impairment 19.6%, Normal/No Impairment 15.0%, Vascular/ Possible Vascular Dementia 8.9%, 6.0% Dementia NOS, Mixed Dementia 5.4%, Lewy Body Dementia 1.1%, Frontotemporal Dementia 0.6%).

Wilde (2006) investigated the factor structure of the RBANS utilizing a sample of 210 patients (50.5% female) who had an ischemic stroke and were completing inpatient rehabilitation. Average age was 61.9 (13.97) years, average education was 12.27 (3.06), and patients were racially and ethnically diverse (Caucasian 59.5%, African American 42%, Hispanic 7.6%, and Asian 1%). Location of stroke varied within this sample (Left 37%, Right 44%, Bilateral 19%), as did lesion location (Cortical 38%, Subcortical 31%, Posterior fossa 16%, Multifocal 15%).

RBANS data from a non-clinical community dwelling elderly group of volunteers, commonly referred to as the Oklahoma group (n = 796), were investigated in a previous factor analytic study (Duff et al., 2006), an age-and-education correction study (Duff et al., 2003) and numerous other RBANS studies (Duff et al., 2009; Duff et al., 2008; Duff et al., 2007; Duff et al., 2005; Patton, Duff, Schoenberg, Mold, Scott, &

Adams, 2005). The correlation matrix provided by Duff had a slightly different sample size than was reported in initial publication. Given this, sample characteristics presented in this research are approximated based on Duff and colleagues published factor analysis. Individuals within this sample were estimated to be on average 73.4 (5.8) years old and primarily Caucasian (86%). There were slightly more women than men (Female 58%). The majority of these participants were cognitively intact and likely to have completed at least high school (59%).

An additional sample of RBANS data was provided by Duff that has not previously been utilized in a factor analytic study. Duff, Hobson, Beglinger, and O'Bryant (2010) investigated the clinical utility of the RBANS in differentiating individuals with Mild Cognitive Impairment (MCI; n = 72) and individuals that are cognitively intact (n = 71). The correlation matrix with RBANS data provided was comprised of a slightly larger sample size than reported in the publication (N = 173) so sample characteristics are again approximates based on previously published material. Average age of the entire sample was approximately 78.7 (7.7) years and mean education was 15.4 (2.5) years (Duff et al., 2010). Individuals were primarily women (81%) and all were Caucasian.

Procedures: Statistical Analysis

PCA was conducted to evaluate the underlying dimensional structure of each sample. While this method technically extracts *components*, the term factor will be used interchangeably since this is common in the literature. As previously described, the goal of PCA is to investigate the correlations between variables (i.e., subtests) and organize this information into a smaller number of factors that infer underlying constructs. The methodological steps and decisions in the current project are presented below.

Tabachnick and Fidell (2013) caution that samples should not be pooled in analyses since they may differ in unknown ways that might impact the underlying factor structure (or cause it to subtly shift) so each sample was investigated individually. Prior to conducting analyses, the Kaiser-Meyer-Olkin (KMO) statistical index was reviewed to evaluate whether there was problematic collinearity between variables (Kaiser, 1981). A KMO statistic greater than .70 indicates that the data is well suited for analysis due to the indication that variance is shared across variables and not only between pairs of variables (Hoelzle & Meyer, 2013). All previously published samples were appropriate for analysis (Carlozzi et. al., 2008, KMO = .91; Duff et al., 2006, KMO = .88; Duff et al., 2010, KMO = .83; Wilde, 2006, KMO = .87). The Vogt sample KMO was .87, which also indicates the data was suitable for PCA.

In PCA the greatest amount of shared variance is identified and assigned to the first factor, the next largest amount of shared variance is brought in line with the second component, and this process continues for subsequent components until all variance is accounted for (Tabachnick & Fidell, 2013). The greatest amount of variance is always extracted in the first component and less in subsequent components, and the amount of variance credited to each is reflected in a standardized eigenvalue (Hoelzle & Meyer, 2013). It is necessary to consider the number of variables present to determine the amount of variance explained by an eigenvalue. In the current study, there are twelve RBANS subtests so if the first component has an eigenvalue of 8.00 it accounts for 66.67% (e.g. 8.00/12 * 100) of the total variance.

A key methodological decision in EFA is determining how many factors to extract from the observed correlation matrix. While extracting too many factors may result in a solution that more closely recreates the original correlation matrix, it increases the odds that meaningful factors will split and result in unreliable components (Fava & Velicer, 1992b). If a parsimonious solution is sought, the investigator may risk extracting too few factors, combining distinct components and oversimplifying the solution (Fava & Velicer, 1992b). Employing empirically supported extraction techniques improves the likelihood that a reliable solution will emerge across diverse samples (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hoelzle & Meyer, 2009; O'Connor, 2002). Supported and unsupported procedures will be presented in the following paragraphs.

A simple procedure often used to guide retention decisions is Kaiser's Criterion, which states that all components with eigenvalues greater than one should be retained (Kaiser, 1960). The problem with this approach is that the number of eigenvalues greater than one is directly related to the number of variables analyzed. The number of components retained typically ranges between one-fifth to one-third of the total number of variables analyzed, regardless of the actual underlying structure of data (Zwick & Velicer, 1982). If Kaiser's criterion were the only extraction utilized in the present study, it might be predicted that two to four components would be expected to have eigenvalues greater than one. Published RBANS factor analytic studies support this prediction, Kaiser's criterion consistently recommended retention of two or three components (see Table 2; Carlozzi et al., 2008; Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010; Wilde, 2006). Despite this method being commonly utilized, empirical research conclusively demonstrates that Kaiser's criterion regularly results in over-extraction and inconsistent component solutions (Fabrigar et al., 1999; Hubbard & Allen, 1987; Preacher & MacCallum, 2003; Zwick & Velicer, 1982).

Visual examination of the scree plot, or eigenvalue plot, is another frequently utilized technique for component extraction (Cattell, 1966). The researcher examines the scree plot to look for the elbow, or sharp break in the curve since the earlier eigenvalues will always be larger than subsequent values. While this approach works well when there are unique factors that account for significant amounts of variance, the technique tends to be highly subjective, so alternative factor extraction or retention guidelines should be utilized as well (Goldberg & Velicer, 2006). When factor differentiation is weak, researchers unreliably identify the sharp break between descending eigenvalues (Linn, 1968; Zwick & Velicer, 1982).

Parallel analysis (PA) also examines eigenvalues, but is considered a more reliable technique since sampling error is considered (Horn, 1965). PA involves generating correlation matrices from random data that includes the same number of variables and subjects as the actual correlation matrix. The eigenvalues from the randomly generated data are then compared to the actual eigenvalues and only factors with eigenvalues greater than those from the random data are retained. Simulated empirical investigations have reported that PA is one of the most accurate methods in determination of the dimensions present in PCA (Crawford, Green, Levy, Lo, Scott, Svetina, Thompson, 2010; Velicer et al., 2000; Zwick & Velicer, 1986).

The Minimum Average Partial (MAP) procedure is an alternative extraction technique initially designed for PCA (Velicer, 1976). The MAP procedure sequentially removes each component from the original correlation matrix and then creates a partial correlation matrix. As each component is removed, the average of the squared partial correlations is computed. As long as each component contains common variance, the average of the squared partial correlations should decrease. This value increases when the component consists of unique variance, and at that point suggests over-extraction. In other words, the suggested number of components to retain is determined at the point at which the average squared partial correlation is smallest. Empirical research has determined that the MAP procedure is the most reliable extraction technique (Zwick & Velicer, 1982, 1986).

In summary, there are a number of different procedures that researchers have followed to determine how many factors should be extracted in PCA. Unfortunately, the methods most often utilized, Kaiser's criterion and the interpretation of scree plots, are most likely to result in non-replicating solutions. Factor retention decisions in the present study are based upon PA and MAP procedure results.

After determining how many factors will be extracted, the next step is to rotate the matrix of loadings to aid interpretability (Golberg & Velicer, 2006). An orthogonal rotation creates a simple structure by producing 90-degree angles between all components so that the correlations between them are zero. In contrast, oblique rotation does not distort relationships between components allowing for the actual relationships between constructs to emerge (Hoelzle & Meyer, 2013). As noted previously, the decision was made to implement oblique rotation for theoretical reasons and because empirical research has demonstrated that cognitive constructs are correlated (Carroll, 1993; Deary, 2000; Hoelzle, Nelson, & Smith, 2010). In the present study an oblique rotation, Direct Oblimin, was utilized. Finally, factor solutions were carefully reviewed to determine what latent constructs have been identified. Interpretatively, items with strong loadings will reflect the cognitive construct whereas those variables with loadings near zero will indicate the absence of a construct.

Utilization of empirically validated methods is likely to result in the most reliable and robust solutions, however, it does not quantify the similarity of solutions obtained from different samples. Often CFA is utilized to determine fit of a solution across samples, however, for reasons previously described (e.g. poor sensitivity to discreet relationships, misfit in large samples) it is not always the most optimal approach. Orthogonal vector matrix comparison (Barrett, 1986) is an alternative method to compare the congruency of multiple factors across samples. Implementation of this technique evaluated the similarity of RBANS factor solutions beyond visual examination of loadings (as previous RBANS factor analytic studies have done). This is an important aspect of this research because solutions can sometimes appear to be inconsistent when they are actually similar. Orthogonal vector matrix comparison methods rotate one sample structure in order to align it with a solution from another sample (Barrett, 1986; Barrett, Petrides, Eysenck, & Eysenck, 1998). Rotation occurs to maximally align the solutions in three dimensional space, without distorting the original component solutions, when a sample solution is compared to a target solution (Barrett, 1986; Barrett et al., 1998; Hoelzle & Meyer, 2009; Hopwood & Donnellan, 2010).

Vector matrix comparison methods result in congruence coefficients that indicate how well factors match one another (Hopwood & Donnellan, 2010). Recommendations for interpreting congruence coefficients vary somewhat. Barrett (1998) suggests benchmarks that are at least .80 to .95 to demonstrate good similarity and coefficients at

.98 and above indicate an identical factor structure between samples. More refined interpretive guidelines have been put forth as well; excellent = .98 – 1.00, good .92 - .98, borderline = .82 - .92, poor = .68 - .82 (MacCallum, Widaman, Zhang, & Hong, 1999). In the present study, orthogonal vector matrix comparisons were completed using Orthosim 2.1 software (Barrett, 2005) to quantify the similarity of RBANS structure across different samples.

Comparison of single component structures requires a different statistical process than multidimensional component structures. Tucker's Congruence Coefficient accounts for both the pattern and magnitude of loadings in order to determine if a single factor solution is replicated across samples (Levine, 1977; Korth & Tucker, 1975). Benchmarks for interpretation of congruence coefficients are reported as; similar = .85 - .94 and identical = .95 – 1.00 (Lorenzo-Seva & ten Berge, 2006). Additionally, single component structures can be compared using Pearson's *r* when a solution has few small loadings (<.40) to, again, compare pattern and magnitude of loadings. Multiple small loadings within a factor will generate a large *r* value masking the impact of more significant loadings, so caution is warranted when utilizing Pearson's *r* (Lorenzo-Seva & ten Berge, 2006). In summary, conclusions regarding replication of invariant structure across diverse samples are based upon vector matrix comparisons, Tucker's congruence coefficient, and Pearson's *r*.

Results

Factor Retention

PCA was conducted separately for each sample. The afore-mentioned factor retention guidelines (e.g. Kaiser's Criterion of Eigenvalues >1, Cattell's visual examination of the Scree Plot, Horn's Parallel Analysis, and Velicer's MAP) were considered and the respective number of components suggested by each are presented in Table 5. Not surprisingly given limitations previously discussed, Kaiser's criterion and visual examination of the Scree Plot resulted in discrepant recommendations regarding how many factors to retain across samples. For example, Kaiser's criterion and visual examination of the Scree Plot suggested retention of one, two, three, or four factors across and within samples. Whereas, PA and MAP indicated retention of either 1 or 2 factors and demonstrated much greater consistency within samples. PA and MAP data analysis procedures are described further in the following paragraphs. Given that these methods are considered superior to others, two and one factor solutions will be further investigated.

Table 5

Principal Components Analysis Extraction Criteria Results Summary

	Carlozzi et al. (2008)	Duff et al. (2010)	Duff et al. (2006)	Wilde (2006)	Vogt
Sample Size	175	173	796	210	345
EV >1	2	4	2	2	3
Scree Plot	1	2	3	2	1
PA	1	1	2	2	2
MAP	1 - 2	1	1 - 2	2	1 - 2

Note: EV = Eigenvalue, PA = Parallel Analysis, MAP = Minimum Partial Average; MAP ranges reflect minor differences between MAP procedures not exceeding .04

PA was conducted individually with each sample using O'Connor's (2000) syntax and results are presented in Table 6. PA compares actual eigenvalues to eigenvalues from 500 randomly generated datasets that have the same parameters as the actual data. In this analysis, 500 correlation matrices of random data were generated with the same number of subtests (e.g. 12) and matched sample size. PA recommends that a component should be retained when the actual eigenvalue is larger than the corresponding randomly generated eigenvalue. Zwick and Velicer (1986) recommends comparing actual eigenvalues to the 95th percentile of randomly generated eigenvalues (as opposed to mean eigenvalue) to decrease risk of over-extraction in situations when sample sizes are small and expected factor loadings are low. In the present study, PA indicated retaining one factor in two samples (Carlozzi et al., 2008; Duff et al., 2010) and two factors in the other three samples (Duff et al., 2006; Wilde, 2006; Vogt). Interpretively, retention recommendations would not have changed if actual eigenvalues were compared to the mean PA eigenvalues as opposed to the 95th percentile of randomly generated eigenvalues.

Table 6

RBANS Actual and Random Eigenvalues from Horn's Parallel Analysis

	Carlozzi et al. (2008)			Duff et al. (2010)			Duff et al. (2006)			Wilde (2006)			Vogt		
	Real EV	*M*	95th EV	Real EV	*M*	95th EV	Real EV	*M*	95th EV	Real EV	*M*	95th EV	Real EV	*M*	95th EV
1	***6.51***	***1.45***	***1.56***	***4.63***	***1.46***	***1.57***	***5.00***	***1.20***	***1.25***	***5.33***	***1.41***	***1.51***	***5.27***	***1.31***	***1.39***
2	1.02	1.32	1.41	1.20	1.33	1.42	***1.35***	***1.15***	***1.19***	***1.98***	***1.30***	***1.37***	***1.38***	***1.23***	***1.28***
3	.89	1.23	1.30	1.07	1.24	1.30	.98	1.11	1.14	.90	1.21	1.27	1.06	1.17	1.21
4	.71	1.15	1.22	1.02	1.15	1.21	.82	1.07	1.10	.69	1.14	1.20	.82	1.11	1.15
5	.58	1.08	1.13	.90	1.08	1.13	.74	1.04	1.07	.65	1.08	1.12	.67	1.06	1.11
6	.47	1.01	1.06	.76	1.02	1.07	.59	1.01	1.03	.50	1.02	1.07	.57	1.01	1.05
7	.41	.95	1.00	.66	.95	1.00	.56	.98	1.00	.46	.96	1.00	.55	.97	1.01
8	.38	.89	.94	.49	.89	.94	.52	.95	.98	.40	.90	.94	.46	.92	.96
9	.35	.83	.88	.44	.82	.88	.49	.92	.95	.35	.84	.89	.39	.88	.92
10	.27	.76	.81	.38	.76	.81	.45	.89	.91	.30	.79	.83	.37	.83	.87
11	.25	.70	.75	.27	.69	.75	.32	.85	.88	.24	.72	.78	.31	.78	.82
12	.17	.62	.68	.18	.61	.68	.19	.81	.84	.21	.64	.71	.16	.72	.77
		1			1			2			2			2	

Note: Real EV = Actual data eigenvalue; *M* = Mean eigenvalue of randomly generated data; 95th EV = 95th percentile eigenvalue of randomly generated data; Bold and italic values indicate the number of components recommended for retention.

Velicer's MAP (1976) procedure was also conducted using syntax generated by O'Connor (2000). In this process, the average squared correlation is computed from each observed correlation matrix. Each component is then partialed out in a compounding fashion (e.g. meaning that in the first step one component is extracted, then in the second step two components are extracted) and the average squared partial correlation is computed at each step. The average squared partial correlation decreases as common variance is continually removed. When an extracted component is based upon unique variance specific to a subtest or pair of subtests, the average partial correlation then increases. So the smallest of the average partial correlations indicated the number of components to extract. Results of the MAP procedure for the present study are shown in Figure 1. In the present study, MAP indicated retaining one factor in Duff et al., 2010 and two factors in a different samples Wilde, 2006. The other three samples (Carlozzi et al.; 2008, Duff et al., 2006; Vogt) exhibited two average partial correlations that were extremely close (e.g. < 0.04) suggesting that both 1 and 2 factor solutions should be explored.

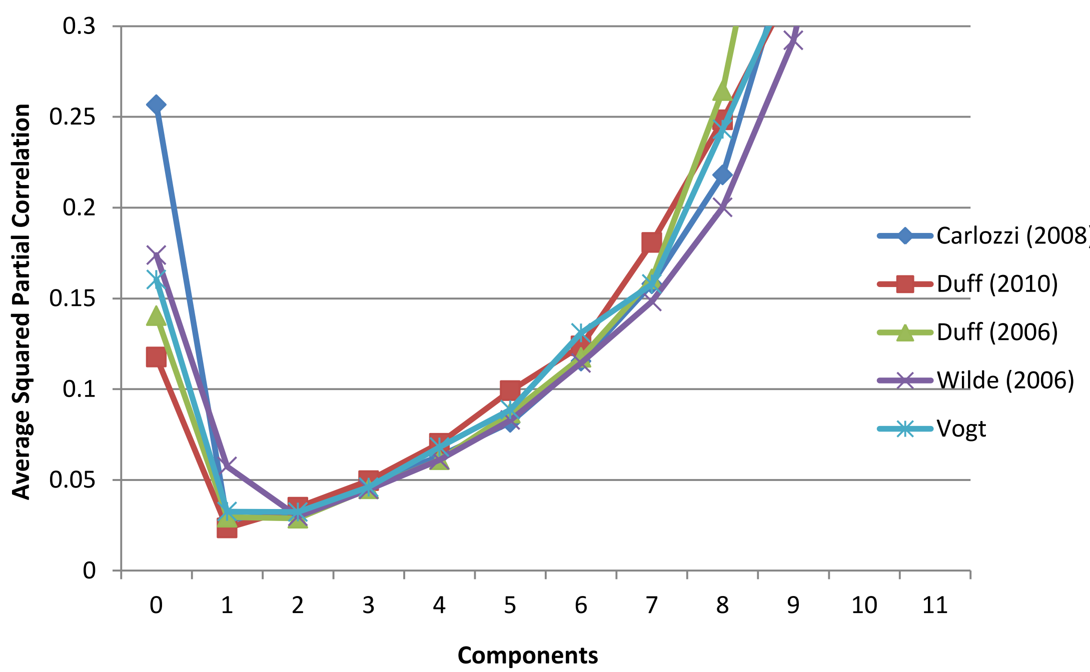

Figure 1. Velicer's MAP procedure indicating number of components to be retained for each RBANS sample.

Component structure

PCA was conducted specifying a two component solution for each sample and solutions were rotated utilizing Direct Oblimin rotation to allow for correlated dimensions. Factor loading results for each sample are presented in Table 7. Latent constructs were inferred by considering the magnitude of factor loadings. Examination of two factor solutions revealed similarity across diverse samples. The first Factor strongly suggests a Memory construct (List Recall, Story Recall, List Learning, List Recognition, Story Memory, Semantic Fluency, and Figure Recall). The Memory factor explains the majority of RBANS score variance (see Table 7; Range 39% to 54% of Total Score variance across samples). Since PCA conducted systematically across samples the factor loadings displayed in Table 7 can be equated and averaged across samples to offer a simplified picture of the factor structure.

It is notable that the Figure Recall subtest displayed meaningful cross loading in two samples, and in the Wilde (2006) sample the subtest is strongly associated with a non-memory dimension. Nevertheless, most reliably, Figure Recall is associated with Factor 1. The latent construct of the first factor is conceptualized as primarily comprised of memory tasks. Semantic Fluency subtest, a verbal fluency task that involves rapidly recalling information from specific categories, also reliably loads there. This verbal fluency task may be conceptualized as a language, executive functioning, or memory task. In this two component solution, it appears the latent construct of memory retrieval emerges to converge with other RBANS memory subtests.

Factor 2 appears to reflect a Visuospatial construct (Figure Copy, Line Orientation, and Coding). The Coding subtest has meaningful factor loadings on both dimensions. These cross loadings could be attributed to the attentional and visuospatial component required in Coding that is conceptually similar to the attention requirements in list and story learning tasks. Additionally, Coding and Semantic Fluency possess a mutual speed component and performance in each of these tasks could be similar. However, Coding loads most reliably onto the second visuospatial component. The second visuospatial factor accounted for between 9% and 17% of the total score variance, which is meaningfully less than the first factor.

Two remaining subtests, Picture Naming and Digit Span did not consistently load on either factor. In the Carlozzi and colleagues sample (2008) and in the Duff and colleagues (2006) sample the picture naming subtest loaded onto the second Visuospatial factor. However, in the Wilde (2006) sample the Picture Naming subtest loaded onto the first Memory factor. When loadings were average across samples, the Picture Naming subtest did not load on to either factor. The Digit Span subtest loaded on the second factor in the Carlozzi and colleagues (2008) samples, however, in the Wilde (2006) sample Digit Span loaded on the first factor. Again, when average loadings were examined across samples the Digit Span subtest loadings were not strong enough to reliably load on either factor.

Table 7

Two Component RBANS Oblique Rotated Pattern Matrices

	Carlozzi et al. (2008)		Duff et al. (2010)		Duff et al. (2006)		Wilde (2006)		Vogt		Average Loadings	
	1	2	1	2	1	2	1	2	1	2	1	2
List Learning	**.75**	.14	**.85**	.06	**.75**	.09	**.86**	-.03	**.76**	.12	**.79**	.08
Story Memory	**.77**	.11	**.72**	.06	**.75**	.11	**.78**	-.09	**.75**	.12	**.75**	.06
Figure Copy	-.07	**.85**	-.22	**.81**	-.05	**.76**	-.15	**.96**	-.12	**.80**	-.12	**.84**
Line Orientation	-.04	**.89**	.07	**.48**	-.16	**.83**	-.04	**.79**	-.01	**.81**	-.04	**.76**
Picture Naming	.20	**.56**	.23	.25	.03	**.62**	**.69**	-.03	.24	.35	.28	.35
Semantic Fluency	**.65**	.14	**.60**	.11	**.40**	.23	**.70**	.08	**.62**	.18	**.59**	.15
Digit Span	.06	**.51**	.12	.38	.09	.37	**.50**	-.02	.23	.14	.20	.28
Coding	**.48**	**.48**	**.59**	.27	.24	**.59**	.31	**.66**	.22	**.70**	.37	**.54**
List Recall	**.94**	-.22	**.87**	-.18	**.89**	-.10	**.74**	.08	**.89**	-.17	**.87**	-.12
List Recognition	**.66**	.16	**.83**	-.28	**.80**	-.12	**.80**	-.01	**.78**	-.08	**.77**	-.07
Story Recall	**.89**	-.02	**.81**	-.20	**.82**	.08	**.77**	.09	**.91**	-.08	**.84**	-.03
Figure Recall	**.45**	**.42**	**.60**	.15	**.36**	**.46**	.10	**.79**	**.74**	.01	**.45**	.37
Eigenvalue	6.51	1.02	4.63	1.20	5.00	1.35	5.33	1.98	5.27	1.38		
Correlation	.61		.53		.48		.36		.44			
Percent of Variance Explained	54.27	8.49	38.54	10.0	41.70	11.25	44.41	16.49	43.92	11.50	44.57	11.6
Total Variance Explained	62.75		48.57		52.96		60.89		55.42		56.12	

Additional analyses were conducted to evaluate one factor RBANS solutions since PA and MAP provided some support for retaining only one factor in several samples (see Table 8). The majority of RBANS subtests meaningfully loaded onto the factor. Subtests with strongest loadings were generally memory tasks indicating the primary presence of the cognitive construct of memory, though language, processing speed, and perceptual organization are also meaningfully emphasized. Digit Span had relatively low loadings on the one factor solution (Pattern matrix loadings < .40) in three samples. The amount of variance explained in the single factor solution mirrors the amount of variance explained by the Memory factor in the two factor solution (see Tables 7 and 8).

Table 8

RBANS Single Component Solution

	Carlozzi et al. (2008)	Duff et al. (2010)	Duff et al. (2006)	Wilde (2006)	Vogt	Averaged Loadings
List Learning	**.82**	**.86**	**.76**	**.80**	**.81**	**.81**
Story Memory	**.82**	**.74**	**.78**	**.68**	**.80**	**.76**
Figure Copy	**.65**	.07	**.55**	**.50**	**.41**	**.44**
Line Orientation	**.71**	.23	**.50**	**.51**	**.52**	**.49**
Picture Naming	**.66**	.32	**.51**	**.63**	**.46**	**.52**
Semantic Fluency	**.72**	**.64**	**.55**	**.71**	**.72**	**.67**
Digit Span	**.49**	.25	.36	**.46**	.31	.37
Coding	**.86**	**.68**	**.67**	**.72**	**.67**	**.72**
List Recall	**.70**	**.80**	**.74**	**.75**	**.75**	**.75**
List Recognition	**.75**	**.73**	**.64**	**.75**	**.70**	**.71**
Story Recall	**.82**	**.80**	**.82**	**.79**	**.83**	**.81**
Figure Recall	**.78**	**.65**	**.70**	**.61**	**.72**	**.69**
Eigenvalue	6.51	4.63	5.00	5.33	5.27	5.35
Percent of Variance Explained	54.27	38.54	41.70	44.41	43.92	44.57

Component comparison

While using the “eyeball test” to look for similarities in patterns and loadings can be informative, it does not provide conclusive evidence of pattern replication (Levine, 1977). Quantitative methods of factor comparison were utilized to determine if an invariant structure replicated across samples. Results of vector matrix comparison of the two component solution using Orthosim 2.1 (Barrett, 2005) are displayed in Table 9. General interpretation of congruence coefficients are based upon two sets of benchmark recommendations. As offered by Barrett and colleagues (1998), congruency coefficients of .85 or greater indicate a replicated factor structure and coefficients of .98 or higher indicate identical solutions. More delineated guidelines offer benchmarks for congruency as; excellent = .98 – 1.00, good .92 - .98, borderline = .82 - .92, poor = .68 - .82 (MacCallum et al., 1999). As stated previously, when vector matrix comparisons are conducted each sample is individually designated as the target sample and then the other samples are sequentially compared to that primary sample. Resulting congruence coefficients vary slightly dependent upon which sample is the primary sample so all congruency coefficients are reported in Table 9.

Overall, vector matrix comparisons strongly support a two component solution with all coefficients except 1 meeting Barrett’s (1998) guidelines for factor replication (see Table 9). In addition when considering the delineated guidelines, 33 out of 40 congruence coefficients meet MacCallum and colleagues (1999) good or excellent benchmarks. Interestingly, there were several instances of borderline congruence in second factor comparisons with the Wilde (2006) sample when compared to Carlozzi et al. (2008) and Duff et al. (2010). The Orthosim program specifies which test variables are involved when there is misfit (i.e., low congruency). Picture Naming and Digit Span subtests displayed poor congruency. When PCA was conducted, the Wilde (2006) sample was the only sample in which the Picture Naming and Digit Span subtests loaded strongly onto the first factor. Further, when PA or MAP recommended retention of a single factor in the Carlozzi and colleagues (2008; PA) and the Duff and colleagues (2010; MAP) samples, there is suggestion of a weaker second factor relative to other samples. These issues likely contributed to the subtly lower congruency coefficients.

Table 9

Two Component Vector Matrix Comparisons with 12 RBANS Subtests

	Carlozzi et al. (2008)		Duff et al. (2010)		Duff et al. (2006)		Wilde (2006)		Vogt	
	1	2	1	2	1	2	1	2	1	2
Carlozzi (2008)	-	-	.96	.94	.95	.93	.95	.88	.98	.95
Duff (2010)	.98	.94	-	-	.97	.97	.93	.87	.98	.94
Duff (2006)	.99	.96	.97	.95	-	-	.95	.90	.99	.96
Wilde (2006)	.94	.91	.95	.84	.93	.92	-	-	.95	.88
Vogt	.98	.96	.98	.92	.98	.96	.95	.90	-	-

In order to investigate the similarity of one component solutions across samples, two methods of single component comparison were utilized. Tucker's Congruence Coefficients were calculated (Levine, 1977; Lorenzo-Seva & ten Berge, 2006) and nearly all samples displayed identical excellent congruence with each other (see Table 10, below the diagonal). Additionally, Pearson *r* correlations were calculated to investigate relationships between the single component structures and loadings (Levine, 1977). All single component solutions were significantly, positively correlated (Table 10; above the diagonal). Both procedures indicate a single factor RBANS dimension is invariant across samples.

Table 10

Single Component Solution Comparisons with 12 RBANS Subtests

	Carlozzi et al. (2008)	Duff et al. (2010)	Duff et al. (2006)	Wilde (2006)	Vogt
Carlozzi (2008)	-	.72**	.84**	.74**	.85**
Duff (2010)	.94	-	.83**	.92**	.93**
Duff (2006)	.99	.96	-	.80**	.91**
Wilde (2006)	.99	.96	.99	-	.88**
Vogt	.99	.98	.99	.99	-

Note: Tucker's Congruence Coefficients located below the diagonal and Pearson's *r* values are above the diagonal. ** $p < .01$

Discussion

Neuropsychological test validation is an ongoing process that requires examination of a measure utilizing multiple clinical and non-clinical samples. Test validity is directly related to clinical utility and thus an important area of focus for researchers and clinicians, alike. The present study investigated the factor analytic structure of the RBANS, a widely used neuropsychological measure (e.g., see Randolph, 1998, 2012). To date, six studies have been conducted to evaluate the RBANS factor structure and reported slightly different solutions (see Tables 2 and 3; Carlozzi et al., 2008; Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010; Wilde, 2006). However, any comparison of previous factor analytic solutions is confounded because different methods were utilized. Many researchers have explained that divergent factor analytic findings are related to sample based differences. However, it seems plausible that solution discrepancies are actually the result of methodological decisions, such as the decision to retain factors with eigenvalues greater than one and the use of orthogonal rotation. Nonetheless, this body of literature clearly suggests that CFA and EFA results are inconsistent with the theoretically developed RBANS five index and single neuropsychological score structure (Carlozzi et al., 2008; Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010; Wilde, 2006).

Other non-factor analytic RBANS studies have also demonstrated poor internal consistency and construct validity of select indices, most notably Attention and Language (Beatty et al., 2003; Beatty et al., 2004; Larson et al., 2005, McKay et al., 2007). While the RBANS is marketed as a stand-alone core battery or screening tool to evaluate multiple cognitive domains (e.g. Immediate Memory, Visuospatial/Construction, Attention, Language, Delayed Memory; Randolph, 2012; 1998), empirical research suggests that clinicians should consider the degree to which the RBANS successfully does this. This study is novel because empirically supported factor retention methods were uniformly applied to multiple samples to identify an invariant RBANS structure and quantitative methods were utilized to evaluate structure replication across samples. The data driven investigation of this measure reveals a strong first component of memory and a second visuospatial component, indicating that the five domain approach may be too far reaching.

Of primary concern in the current investigation was inclusion of empirically validated methods in order to determine the most reliable factor solution. Application of consistent extraction method (e.g. PCA) and rotation (e.g. Oblique) allowed similar factor structures to emerge across solutions. Factor retention decisions, however, are arguably the most critical to structure conclusions (Hayton et al., 2004; Hoelzle & Meyer, 2009; Hubbard & Allen, 1987; O'Connor, 2002; Zwick & Velicer, 1982). Consistent with expectations, Kaiser's criterion and visual examination of the scree plot displayed inconsistency in factor retention recommendations both across and within samples. Horn's PA (1965) and Velicer's MAP (1976) procedures indicated retention of one or two factors. One factor retention were suggested from PA and MAP in the Carlozzi and colleagues (2006; PA) and Duff and colleagues (2010; MAP) samples, whereas, PA and MAP suggested retention of two factors in the remaining samples. Hence, both two and one factor solutions were explored to alleviate risk of over- or under-extraction. Under-extraction creates loss of important information and neglect of potentially important latent constructs, whereas, over-extraction diffuses data and places too much importance on trivial factors (Fava & Velicer, 1992b; Hayton et al., 2004; Wood et al.,1996). Thus, balance is important and was carefully examined.

Previous researchers have purported that differences in the number of factors to retain, the pattern in which subtests load onto factors, and the emergence of latent constructs could be sample specific (Delis et al., 2003; Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010; Wilde, 2006). Analysis of diverse samples, both clinical and non-clinical, revealed that a replicable solution does in fact emerge. Previous literature has utilized an “eye ball” method to infer similarity of a two factor RBANS solution across samples (Carlozzi et al., 2008; Duff et al, 2009; Garcia et al., 2008). The present study was the first to utilize a quantitative method to evaluate solutions of multiple samples. Vector matrix comparison revealed the presence of an invariant two factor RBANS solution across diverse samples (e.g. see Barrett, 2005). Furthermore, this invariant structure demonstrates that factor analytic solution discrepancies that appeared in the literature previously are not due solely to sample characteristics but rather methodological decisions.

Utilizing PCA, a two factor RBANS solution clearly emerges across multiple samples with a first prominent memory factor and second visuospatial factor. Furthermore, the majority of congruency coefficients were good to excellent in vector matrix comparisons. Interestingly, two subtests, Picture Naming and Digit Span, did not consistently load on either factor when investigating pattern matrix loadings (see Table 7), and this minimally impacted overall congruency of solutions because the loadings were not prominent in defining factors. Notably, in several comparisons, these subtests did contribute to slightly lower congruency coefficients between two respective samples. Additional exploratory analyses were performed to evaluate the replication of a two factor solution with Picture Naming and Digit Span removed. Overall, this improved the majority of congruence coefficients (see Table 11) and confirms that these subtests contributed to lower than exceptional congruency across samples.

Table 11

Two Component Vector Matrix Comparisons with 10 RBANS Subtests

(Picture Naming and Digit Span Removed)

	Carlozzi et al. (2008)		Duff et al. (2010)		Duff et al. (2006)		Wilde (2006)		Vogt	
	1	2	1	2	1	2	1	2	1	2
Carlozzi (2008)	-	-	.97	.92	.98	.98	.98	.96	.98	.95
Duff (2010)	.97	.96	-	-	.97	.98	.97	.94	.98	.93
Duff (2006)	.98	.98	.98	.95	-	-	.98	.98	.99	.96
Wilde (2006)	.98	.95	.98	.91	.98	.98	-	-	.97	.90
Vogt	.98	.96	.98	.90	.98	.97	.97	.91	-	-

Given that PA and the MAP procedure provided some support for the retention of one factor in select samples (see Table 5), PCA was again conducted and one factor solutions were investigated. Examination of this solution revealed that the majority of subtests loaded onto the single component with the exception of Digit Span in most of the samples. Additionally, quantitative analysis of factor congruency across samples revealed strong evidence for solution replication. However, when a one factor solution was specified, the amount of variance explained mirrored the first memory factor. In addition, the subtests that most strongly defined the dimension were tasks involving memory. It appears that the underlying cognitive construct of single total score of the RBANS is not general neuropsychological status, but rather predominantly memory functioning. This suggests that an empirically derived single factor score would be most sensitive to memory deficits as opposed to other cognitive issues that a patient might be experiencing.

A single factor structure, in comparison to the robust two factor solution, compresses RBANS subtests that evaluate visuospatial functioning. This is a clear disadvantage for the one factor solution. Moreover, one could argue that, clinically a two factor solution is more informative and would have greater clinical utility.

As an exercise to demonstrate the potential drawback of over-extraction, additional analyses were conducted to explore a three factor RBANS solution. One previous study reported a three factor solution (Garcia et al, 2008) and two extraction criteria (e.g. Kaiser's criteria and visual examination of the scree plot) indicated the possibility of retaining three factors. A three factor solution clearly resulted in over-extraction since the third factor was typically only defined by Digit Span and the other RBANS subtests shifted between factors 1 and 2 in an inconsistent manner (see Appendix A). Additionally, vector matrix comparisons indicated poor replication of a three factor solution (see Appendix B). Specifically, many of the congruency coefficients were in the borderline, or lower, range (26/60) and with only a few coefficients in the exceptional range (8/60). These findings clearly demonstrate the importance of utilizing empirically supported factor retention strategies (e.g. PA and MAP) in order to identify an invariant factor structure.

Importantly, the present study revealed valuable information regarding specific indices and subtests within the RBANS. Attention and Language indices did not emerge in this study nor in previous factor analytic studies (Carlozzi et al., 2008; Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010; Wilde, 2006) due to typically low (or at best moderate) relationships between subtests that comprise these indices. Previous RBANS literature has revealed poor internal consistency of the

Attention and Language indices (Beatty et al., 2003; Beatty et al., 2004; Larson et al., 2005, McKay et al., 2007) and the current research offers further indication poor construct validity. Empirical investigation reveals that often a minimum of three measures assessing a common construct must be present for a related component to emerge (Velicer & Fava, 1998). The RBANS does not possess enough purely language and attention tasks for these indices to emerge in factor analytic investigations.

It is not surprising that Picture Naming and Digit Span subtests do not appear to reliably load with Memory or Visuospatial factors given the discrepancy between constructs. Examination of the individual correlation matrices reveals small associations between these two subtests with other RBANS subtests. There is a ceiling effect (i.e., concentration of scores at the top range with small variance) present in the Picture Naming subtest in these samples (see Appendix C; Range of M = 8.87 to 9.56; Range of SD = 0.81 to 1.53) and in the normative sample (Picture Naming M = 9.47 SD = 0.73; Randolph, 1998). Restricted range in a subtest attenuates the relationships between that task and others within the test (Fabrigar et al., 1999). In other words, a skewed subtest is limited in its ability to meaningfully correlate with other subtests that are more normally distributed (Goldberg & Velicer, 2006). Additionally, a ceiling effect present in both clinical (Carlozzi et al., 2008; Wilde, 2006; Vogt sample) and non-clinical samples (Duff et al., 2006; Randolph, 1998) suggests potentially limited clinical utility of the Picture Naming subtest. Clinical implications of this finding might involve either a revision of the Picture Naming subtest to include more items and increase the difficulty of confrontation naming items or consideration could be given to eliminating the subtest in an RBANS revision.

Examination of the RBANS two component solution reveals low loadings for the Digit Span subtest and inconsistency in loading on either factor. The RBANS digit span forward task is conceptually an attention task. As stated previously, empirical investigation reveals that often a minimum of three measures assessing a common construct must be present for a related component to emerge (Velicer & Fava, 1998). Consequently, there simply is not sufficient representation of this construct to enable Digit Span to load reliably onto a component. Interpretation of single subtest to represent a cognitive construct may not be optimally reliable nor sensitive and may ultimately impact the clinical utility of the measure. Further, standardized testing procedure dictate administration of only the first trial in a set when the first item in the set is passed in interest of brevity. Anecdotally, in clinical settings tasks assessing working memory are frequently administered in addition to the RBANS. Recommended revision to the RBANS could include expansion of the digit span task to include backward and sequencing components (similar to the Wechsler Adult Intelligence Scale-Fourth Edition Digit Span subtest; Wechsler, 2009). To further develop a working memory component, an additional working memory task, such as mental arithmetic or letter-number sequencing could also be added to the RBANS. Assessment of working memory could improve clinical utility of the RBANS across diverse populations, as this construct is often impaired in psychiatric (e.g. anxiety and mood disorders) and neurologic conditions (e.g., dementias, mild traumatic brain injury).

Conclusions

Widespread agreement exists that a viable and defensible factor structure does not emerge from a single analysis. An optimal factor structure is one that is replicated across multiple diverse samples, with varying sample size (Goldberg & Velicer, 2006). Exploratory factor analysis can be used to identify whether an invariant structure emerges across samples. The present study has documented an invariant two component solution through exploratory analysis and confirmed pattern replication through vector matrix comparison (Barrett, 2005). These factors primarily reflect Memory (e.g. List Recall, Story recall, List Learning, List Recognition, Story Memory, Semantic Fluency, & Figure Recall) and Visuospatial (e.g. Figure Copy, Line Orientation, & Coding) cognitive constructs within the RBANS. Furthermore, Picture Naming and Digit Span subtests were demonstrated to have low convergence with other RBANS subtests, do not consistently load onto factors, and adversely impact the component replication.

Additionally, the present study has empirically supported the position that differences in RBANS factor solutions are primarily due to methodological decisions and are not solely related to unique sample characteristics. Simply put, the RBANS factor structure is relatively invariant across diverse samples. Previous studies RBANS factor analytic studies (Duff et al., 2006; Garcia et al., 2008; King et al., 2012; Schmitt et al., 2010; Wilde, 2006) reported differences between solutions are due to sample differences, frequently citing Delis and colleagues (2003) investigation of the CVLT. However, O'Connor (2002) empirically demonstrated that invariant solutions can be found in personality measures across diverse samples. This research offers evidence that previously published RBANS solution discrepancies were due to methodological decisions, most importantly, factor retention strategies. The present study uniformly utilized PCA, PA and the MAP procedure to guide factor retention decisions, and oblique rotation. Moreover, empirical methods evaluating replication of factor solutions (Barrett, 2005; Levine, 1977; Lorenzo-Seva & ten Berge, 2006) quantified invariance, which is preferable to the commonly used "eye ball" test. Of note, the present study demonstrated the utility of congruency comparison in a neuropsychological measure. A two component solution reliably emerged and demonstrated good congruence across diverse samples. Also importantly, an invariant structure of the RBANS is apparent across clinical and non-clinical samples.

This investigation of the RBANS provides important clinical insights. The underlying structure of the RBANS suggests the five domain theoretical design of the RBANS is inconsistent with how subtests naturally co-vary. The RBANS component structure suggests Memory and Visuospatial constructs are most reliably assessed. Furthermore, the Picture Naming subtest demonstrates a ceiling effect in clinical and non-clinical samples, thus impacting overall clinical utility. Also, noteworthy the Digit Span subtest does not converge with other tasks within the RBANS. This information in combination with findings that select Index scores have problematic reliabilities, suggests that clinicians should be cautious when interpreting those composite scores.

Future Directions

An identified invariant RBANS factor structure has implications for future research and clinical practice. Component scores could be developed using a normative sample and a unit-weighting scheme or exact factor score approach (Grice, 2001a; 2001b). Each of these analysis procedures could be explored to determine whether the empirically derived factor scores or theoretically developed Index scores were more useful in detecting cognitive impairment or meaningful change from a baseline level of functioning. Factor scores may provide better clinical utility because theoretically, they should have greater reliability and therefore, be more sensitive to change. Duff and colleagues (2009) recognized the likely presence of a two factor solution and developed data for a Verbal and Visual Indices and a Total Scale Index based upon data for the OKLAHOMA sample. The present study strongly supports consideration of Memory and Visuospatial Indices, but raises questions regarding the utility of a Total Scale Index because it would primarily reflect memory functioning. Future exploration of Memory and Visuospatial component scores is warranted in clinical samples with well-defined impairment affecting the respective constructs.

Additionally, findings suggest that future revisions to the RBANS may include revision or elimination of the Picture Naming subtest. The Digit Span subtest could be expanded (e.g. backward and sequencing trials added) and another conceptually similar subtest could be added in order to increase the likelihood that a factor reflecting working memory reliably emerges across samples. A re-conceptualization and revision to the RBANS to allow clinicians the ability to assess verbal working memory within the RBANS may be useful across multiple populations and improve clinical utility.

References

American Academy of Clinical Neuropsychology Board of Directors. (2007) American Academy of Clinical Neuropsychology (AACN) Practice Guidelines for Neuropsychological Assessment and Consultation, *The Clinical Neuropsychologist, 21*, 209-231, doi:10.1080/13825580601025932

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060-1073.

Aupperle, R. L., Beatty, W. W., Shelton, F. N., & Gontkovsky, S. T. (2002). Three screening batteries to detect cognitive impairment in multiple sclerosis. *Multiple Sclerosis, 8*, 382-389. doi:10.1191/1352458502ms832oa

Barrett, P. T. (1986). Factor comparison: An examination of three methods. *Personality and Individual Differences. 7*(3), 327-340. doi:10.1016/j.bbr.2011.03.031

Barrett, P. T. (2005). *Orthosim 2.1 (version 2.1): Target comparison matrix fitting* (Computer software and manual). Retrieved July 1, 2013, from http://www.pbarrett.net/software.

Barrett, P. T., Petride, K. V., Eysenck, S. B. G., & Eysenck, H. J. (1998). The Eysenck Personality Questionnaire: An examination of the factorial similarity of P, E, N, and L across 34 countries. *Personality and Individual Differences, 25*, 805-819. doi:10.1016/j.bbr.2011.03.031

Beatty, W. W. (2004). RBANS analysis of verbal memory in Multiple Sclerosis. *Archives of Clinical Neuropsychology, 19,* 825-834. doi:10.1016/j.acn.2003.12.001

Beatty, W. W., Ryder, K. A., Gontkovsky, S. T., Scott, J. G., McSwan, K. L., & Bharucha, K. J. (2003). Analyzing the subcortical dementia syndrome of Parkinson's disease using the RBANS. *Archives of Clinical Neuropsychology, 18,* 509-520. doi:10.1016/j.bbr.2011.03.031

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-606. doi:10.1037/0033-2909.88.3.588

Benton, A. L., Hamsher, K. de S., & Sivan, A. B. (1994). Multilingual aphasia examination. Iowa City, IA: AJA Associates.

Benton, A. L., Hamsher, K. de S., Varney, N. R., & Spreen, O. (1983). *Judgment of Line Orientation.* New York, NY: Oxford University Press.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31*, 141-154. doi:10.1037//0735-7028.31.2.I41

Carlozzi, N. E., Horner, M. D., Yang, C., & Tilley, B. C. (2008). Factor analysis of the Repeatable Battery for the Assessment of Neuropsychological Status. *Applied Neuropsychology, 15,* 274-279. doi:10.1080/09084280802325124

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-267. doi:10.1207/s15327906mbr0102_10

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York, NY: Plenum.

Chianetta, J. M., Lefebvre, M., LeBlanc, R., & Grignon, S. (2008). Comparative psychometric properties of the BACS and RBANS in patients with schizophrenia and schizoaffective disorder. *Schizophrenia Research, 105,* 86-94. doi:10.1016/j.schres.2008.05.024

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology, 66,* 93-114. doi:10.1037/0022-3514.66.1.93

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Crawford, A. V., Green, S. B., Lo, W. J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement, 70*(6), 885-901. doi:10.1177/0013164410379332

Crum, R. M., Anthony, J. C., Bassett, S. S., & Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. The Journal of the American Medical Association, *269*, 2386–2391. doi:10.1001/jama.1993.03500180078038

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. Oxford, UK: Oxford University Press.

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: 10.10170S1355617703960139

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1994). *The California Verbal Learning Test-Children's Version.* San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *The California Verbal Learning Test* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Dowling, N. M., Hermann, B., La Rue, A., and Sager, M. A. (2010). Latent structure and factorial invariance of a neuropsychological test battery for the study of preclinical Alzheimer's Disease. *Neuropsychology*, *24*, 742-756. doi:10.10.37/a0020176.

Duff, K. D., Beglinger, L. J., Schoenberg, M. R., Patton, D. E., Mold, J., Scott, J. G., & Adams, R.L. (2005). Test-retest stability and practice effects of the RBANS in a community dwelling elderly sample. *Journal of Clinical and Experimental Neuropsychology*, *27*, 565-575. doi:10.1080/13803390490918363

Duff, K.D., Hobson, V. L., Beglinger, L. J., & O'Bryant, S. E. (2010). Diagnostic accuracy of the RBANS in mild cognitive impairment: limitations on assessing milder impairments. *Archives of Clinical Neuropsychology, 25*(5), 429-441.

Duff, K. D., Langbehn, D. R., Schoenberg, M. R., Moser, D. J., Baade, L. E., Mold, J. W., . . . Adams, R. L. (2006). Examining the Repeatable Battery for the

Assessment of Neuropsychological Status: Factor analytic studies in an elderly sample. *The American Journal of Geriatric Psychiatry, 14*, 976-979. doi:10.1097/01.JGP.0000229690.70011

Duff, K. D., Langbehn, D. R., Schoenberg, M. R., Moser, D. J., Baade, L. E., Mold, J. W., ... Adams, R. L. (2009). Normative data on and psychometric properties of verbal and visual indexes of the RBANS in older adults. *The Clinical Neuropsychologist, 23,* 39-50. doi:10.1080/13854040701861391

Duff, K. D., Leber, W. R., Patton, D. E., Schoenberg, M. R., Mold, J. W., Scott, J. G., & Adams, R. L. (2007). Modified scoring criteria for the RBANS figures. *Applied Neuropsychology, 14*, 73-83. doi:10.1080/09084280701319805

Duff, K. D., Patton, D., Schoenberg, M. R., Mold, J., Scott, J. G., & Adams, R. L. (2003). Age-and education-corrected independent normative data for the RBANS in a community dwelling elderly sample. *The Clinical Neuropsychologist, 17*, 351-366. doi:10.1076/clin.17.3.351.18082

Duff, K. D., Schoenberg, M. R., Patton, D. E., Mold, J. W., Scott, J. G., & Adams, R. L. (2008). Predicting Cognitive Change across 3 Years in Community-Dwelling Elders. *The Clinical Neuropsychologist*, *22*, 651-661. doi:10.1080/13854040701448785

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299. doi:10.1037/1082-989X.4.3.272

Fava, J. L., & Velicer, W. F. (1992a). An empirical comparison of factor, image, component, and scale scores. *Multivariate Behavioral Research, 27*, 301–322. doi:10.1207/s15327906mbr2703_1

Fava, J. L., & Velicer, W. F. (1992b). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research, 27*, 387–415. doi:10.1207/s15327906mbr2703_5

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. doi:10.1016/j.bbr.2011.03.031

Gignac, G. E., Bates, T. C., & Jang, K. (2007). Implications relevant to CFA model misfit, reliability, and the five factor model as measured by the NEO-FFI. *Personality and Individual Differences*, *43*, 1051-1062. doi:10.1016/j.paid.2007.02.024

Gold, J. M., Queern, C., Iannone, V. N., & Buchanan, R. W. (1999). Repeatable battery for the assessment of neuropsychological status as a screening test in schizophrenia I: Sensitivity, reliability, and validity. *American Journal of Psychiatry, 156*(12), 1944-1950.

Goldberg, L. R., & Velicer, W. F. (2006). Principles of exploratory factor analysis. In S. Strack (Ed.), *Differentiating normal and abnormal personality* (2nd ed.). (pp. 209-237). New York, NY: Springer.

Gontovsky, S. T., Beatty, W. W., & Mold, J. W. (2004). Repeatable Battery for the Assessment of Neuropsychological Status in a normal, geriatric sample. *Clinical Gerontologist*, *273*, 79-86. doi:10.1300/J018v27n03_07

Goodglass, H. & Kaplan, E. (1983). *Boston diagnostic aphasia examination.* Philadelphia, PA: Williams & Wilkins.

Gorsuch, R. L. (1983). *Factor Analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Grice, J. W. (2001a). A comparison of factor scores under conditions of factor obliquity. *Psychological Methods, 6,* 67-83. doi: 10.1037/1082-989X.6.1.67

Grice, J. W. (2001b). Computing and evaluating factor scores. *Psychological Methods, 6,* 430-450. doi:10.1037//1082-989X.6.4.430

Groth-Marnat, G. (2009). *Handbook of psychological assessment.* Hoboken, NJ: John Wiley & Sons, Inc.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York, NY: McGraw-Hill.

Hayton, J. C., Allen, D. G., Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial in parallel analysis. *Organizational Research Methods, 7,* 191-205. doi: 10.1177/1094428104263675

Hobart, M. P., Goldberg, R., Bartko, J. J., & Gold, J. M. (1999). Repeatable battery for the assessment of neuropsychological status as a screening test in schizophrenia, II: Convergent/discriminant validity and diagnostic group comparisons. *American Journal of Psychiatry, 156*(12), 1951-1957.

Hoelzle, J. B. (2008). Neuropsychological Assessment in the Cattell-Horn-Carroll (CHC) Cognitive Abilities Model. (Unpublished doctoral dissertation). The University of Toledo, Toledo, OH.

Hoelzle, J. B., & Meyer, G. J. (2009). The invariant component structure of the Personality Assessment Inventory (PAI) full scales. *Journal of Personality Assessment, 91*, 175-186. doi:10.1080/00223890802634316

Hoelzle, J. B., & Meyer, G. J. (2013). Exploratory factor analysis: Basics and beyond. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (2nd ed.), pp. 164-188. Hoboken, NJ: John Wiley & Sons, Inc.

Hoelzle, J. B., Nelson, N. W., & Smith, C. A. (2010). Comparison of the Wechsler Memory Scale-Fourth Edition (WMS-IV) and Third Edition (WMS-III) dimensional structures: Improved ability to evaluate auditory and visual constructs. *Journal of Clinical and Experimental Neuropsychology, 33*(3), 283-291. doi:10.1080/13803395.2010.511603

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 19*, 149-161. doi:10.1007/BF02289447

Hubbard, R. & Allen, S. J. (1987). An empirical comparison of alternative methods for principal component extraction. *Journal of Business Research, 15,* 173-190. doi: 10.1016/0148-2963(84)90047-X

Iverson G.L., Brooks B.L., & Haley G.M. (2009). Interpretation of the RBANS in inpatient psychiatry: clinical normative data and prevalence of low scores for patients with schizophrenia. *Applied Neuropsychology*, *16*, 31-41. doi: 10.1080/09084280802644128

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151. doi:10.1077/01316446002000116

Kaiser, H. F. (1981). A revised measure of sampling adequacy for factor-analytic data matrices. *Educational and Psychological Measurement*, *41*, 379-381. doi:10.1177/001316448104100216

Kaplan, E. F., Goodglass, H., & Weintraub, S. (2001). *The Boston Naming Test* (2nd ed.). Philadelphia: Lippincott Williams &Wilkins.

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*, *27,* 756-765. doi: 10.1093/arclin/acs062

Kline, P. (1979). *Psychometrics and psychology*. London, UK: Academic Press.

Korth, B. & Tucker, L. R. (1975). The distribution of chance congruence coefficients from simulated data, *Psychometrika, 40*, 361-372. doi:10.1007/BF02291763

Larrabee, G. J. (2003). Lessons on measuring construct validity: A commentary on Delis, Jacobson, Bondi, Hamilton, and Salmon. *Journal of the International Neuropsychological Society, 9*, 947-953. doi:10.10170S1355617703960140

Larson, E. B., Kirschner, K., Bode, R. K., Heinemann, A. W., Clorfene, J., & Goodman, R. (2003). Brief cognitive assessment and prediction of functional outcome in stroke. *Top Stroke Rehabilitation, 9*, 10-21. doi:10.1310/84YN-Y640-8UEQ-PDNV

Larson, E. B., Kirschner, K., Bode, R. K., Heinemann, A., & Goodman, R. (2005). Construct and predictive validity of the repeatable battery for the assessment of neuropsychological status in the evaluation of stroke patients. *Journal of Clinical and Experimental Neuropsychology, 27*, 16-32. doi:10.1080/138033990513564

Lee, K., & Ashton, M. C. (2007). Factor analysis in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 424-443). New York: Guilford.

Levine, M. S. (1977). *Canonical Analysis and Factor Comparison*. Beverly Hills, CA: Sage Publications, Inc.

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

Linn, R. L. (1968). A monte carlo approach to the number of factors problem. *Psychometrika, 33*(1), 37-71.

Lorenzo-Seva, U. & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2,* 57-64. doi:10.1027/1614-2241.2.2.57

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84-99. doi:10.1037/1082-989X.4.1.84

McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 70,* 552-566. doi:10.1037/0022-3514.70.3.552

McKay, C., Casey, J. E., Wertheimer, J., & Fichtenberg, N. L. (2007). Reliability and validity of the RBANS in a traumatic brain injured sample. *Archives of Clinical Neuropsychology, 22,* 228-241. doi:10.1016/j.bbr.2011.03.031

Messick, S. (1995).Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into scoring meaning. *American Psychologist, 9,* 741–749. doi:10.1037/0003-066X.50.9.741

Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavioral Research Methods, Instruments & Computers, 32*(3), 396-402.

O'Connor, B. P. (2002). The search for dimensional structure differences between normality and abnormality: A statistical review of published data on personality and psychopathology. *Journal of Personality and Social Psychology. 83,* 962-982. doi:10.1037/0022-3514.83.4.962

Pachet, A. K. (2007). Construct validity of the Repeatable Battery of Neuropsychological Status (RBANS) with acquired brain injury patients. *The Clinical Neuropsychologist*, *21*, 286-293, doi:10.1080/13854040500376823

Patton, D. E., Duff, K., Schoenberg, M. R., Mold, J., Scott, J. G., & Adams, R. L. (2005). Base rates of longitudinal RBANS discrepancies at one- and two-year intervals in community-dwelling older adults. *The Clinical Neuropsychologist, 19*, 27-44. doi:10.1080/13854040490888477

Preacher, K. J. & MacCullum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*, 13-43. doi:10.1207/S15328031US0201_02

Psychological Corporation. (2001). *Manual for the Wechsler Test of Adult Reading (WTAR).* San Antonio, TX: Psychological Corporation.

Randolph, C. (2012). *Repeatable battery for the assessment of neuropsychological status update (RBANS Update).* Bloomington, MN: NSC Pearson, Inc.

Randolph, C. (1998). *Repeatable battery for the assessment of neuropsychological status.* Bloomington, MN: NSC Pearson, Inc.

Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary clinical validity. *Journal of Clinical and Experimental Neuropsychology, 20*, 310-319. doi:10.1076/jcen.20.3.310.823

Schmitt, A. L., Livingston, R. B., Smernoff, E. N., Reese, E. M., Hafer, D. G., & Harris, J. B. (2010). Factor analysis of the Repeatable Battery for the Assessment of Neuropsychology Status (RBANS) in a large sample of patients suspected of dementia. *Applied Neuropsychology, 17,* 8-17. doi:10.1080/09084280903297719

Schoenberg, M. R., & Scott, J. G. (2011). The neuropsychology referral and answering the referral question. *The little black book of neuropsychology: A syndrome-based approach*. New York, NY: Springer.

Slick, D. (2006). Psychometrics in neuropsychological assessment. In E. Strauss, E. M. Sherman, & O. Spreen (Eds.), *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York, NY: Oxford University Press.

SPSS Inc. (2011). *SPSS base 20.0 for windows user's guide*. SPSS Inc., Chicago IL.

Stern, R. A., & White, T. (2003). *Neuropsychological Assessment Battery: Administration, Scoring, and Interpretation manual.* Lutz, FL: Psychological Assessment Resources.

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

Sweet, J. J., Peck, E. A., Abramowitz, C., & Etzweiler, S. (2002). National Academy of Neuropsychology/Division 40 of the American Psychological Association Practice Survey of clinical neuropsychology, part 1: Practitioner and practice characteristics, professional activities, and time requirements. *The Clinical Neuropsychologist*, *16,* 109–127. doi:10.1076/clin.16.2.109.13237

Tabachnick, B. G., & Fidell, L. S. (2013). Principal components and factor analysis. *Using multivariate statistics* (6th ed.) pp. 612-680. Boston, MA: Pearson Education, Inc.

Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement*, *58*, 541-568. doi:10.1177/0013164498058004001

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321-327. doi:10.1007/BF02293557

Velicer, W. F. & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3*(2), 231-251. doi:10.1037/1082-989X.3.2.231

Wechsler, D. (2009). *Advanced Clinical Solutions (ACS) for the Wechsler Adult Intelligence Scale -Fourth Edition and Wechsler Memory Scale – Fourth Edition.* San Antonio, TX: Pearson Assessments.

Wilde, M. C. (2006). The validity of the Repeatable Battery of Neuropsychological Status in acute stroke. *The Clinical Neuropsychologist, 20,* 702-715. doi:10.1080/13854040500246901

Wilk, C. M., Gold, J. M., Bartko, J. J., Dickerson, F., Fenton, W. S., Knable, M., . . . Buchanan, R. W.(2002). Test-retest stability of the Repeatable Battery for the Assessment of Neuropsychological Status in schizophrenia. *American Journal of Psychiatry, 159*, 838-844. doi:10.1176/appi.ajp.159.5.838

Wilk, C. M., Gold, J. M., Humber, K., Dickerson, F., Fenton, W. S., & Buchanan, R. W. (2004). Brief cognitive assessment in schizophrenia: normative data for the Repeatable Battery for the Assessment of Neuropsychological Status. *Schizophrenia Research, 70*, 175-186. doi:10.1016/j.schres.2003.10.009

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of over- and under-extraction on principal factor analysis with varimax rotation. *Psychological Methods, 1*(4), 354-365

Woodhouse, J., Hevanka, D. J., Scott, J., Vincent, A., Roebuck-Spencer, T., Domboski-Davidson, K.,... Adams, R. (2013). Efficacy of the ANAM General Neuropsychological Screening battery (ANAM GNS) for detecting neurocognitive impairment in a mixed clinical sample. *The Clinical Neuropsychologist, 27,* 376-385. doi:10.1080/13854046.2012.762427

Zoski, K. W. & Jurs, S. (1996). An objective counterpart to the visual scree test for factor analysis: The standard error scree. *Educational and Psychological Measurement, 56*, 443-451. doi: 10.1177/0013164496056003006

Zwick, W. R. & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research*, 17, 253-269. doi:10.1207/s15327906mbr1702_5

Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules to determine the number of components to retain. *Psychological Bulletin, 99*, 432-444. doi:10.1037/0033-2909.99.3.432

Appendix A

Three Component RBANS Oblique Rotated Pattern Matrices

	Carlozzi et al. (2008)			Duff et al. (2010)			Duff et al. (2006)			Wilde (2006)			Vogt		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
List Learning	**.73**	-.08	.17	**.85**	.06	.02	**.75**	.05	.16	**.86**	-.03	.05	**.65**	.10	.33
Story Memory	**.75**	.11	.05	**.70**	-.03	.15	**.75**	.03	.28	**.56**	-.05	**.43**	**.67**	.12	.20
Figure Copy	-.07	**.90**	-.03	-.13	**.90**	.12	-.05	**.78**	.01	-.13	**.95**	-.01	-.08	**.84**	-.18
Line Orient.	-.04	**.84**	.14	.02	.14	**.60**	-.15	**.77**	.20	-.10	**.85**	.13	-.04	**.82**	.05
Picture Naming	.19	**.61**	-.04	.37	.37	-.09	.03	**.66**	-.07	**.54**	-.01	.28	.23	.37	.01
Semantic Fluency	**.63**	.02	.27	**.59**	.02	.16	**.40**	.26	-.04	**.76**	.06	-.08	**.51**	.15	.32
Digit Span	.05	.05	**.93**	.03	-.13	**.80**	.12	.04	**.91**	.01	.06	**.88**	-.05	-.03	**.92**
Coding	**.47**	**.42**	.19	.57	.09	.32	.23	.61	-.03	.28	**.66**	.07	.12	.67	.26
List Recall	**.92**	-.14	-.11	**.88**	-.09	-.17	**.88**	-.04	-.10	**.79**	.07	-.05	**.88**	-.12	.02
List Recog.	**.65**	.30	-.23	**.78**	-.36	.04	**.79**	-.04	-.18	.90	-.03	-.15	**.79**	-.02	-.07
Story Recall	**.87**	-.02	.06	**.83**	.05	-.08	**.81**	-.04	-.17	**.67**	.11	.21	**.92**	-.01	-.07
Figure Recall	**.44**	**.43**	.04	**.64**	.23	-.06	.35	**.53**	-.14	.21	**.76**	-.18	**.79**	.08	-.20
EV	6.51	1.02	.89	4.63	1.20	1.07	5.00	1.35	.98	5.33	1.98	.90	5.27	1.38	1.06
	—			—			—			—			—		
r	.58	—		.14	—		.46	—		.34	—		.42	—	
	.28	.27	—	.19	.04	—	.09	.17	—	.34	.12	—	.26	.18	—
Percent of Variance Explained	54.3	8.5	7.4	38.5	10.0	8.91	41.7	11.3	8.20	44.4	16.5	7.51	44.0	11.5	8.9

Appendix B

Three Component Vector Matrix Comparison with 12 RBANS Subtests

	Carlozzi et al. (2008)			Duff et al. (2010)			Duff et al. (2006)			Wilde (2006)			Vogt		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Carlozzi (2008)	-	-	-	.96	.91	.75	.98	.97	.90	.96	.90	.85	.98	.95	.95
Duff (2010)	.96	.62	.75	-	-	-	.97	.93	.79	.95	.87	.72	.98	.91	.84
Duff (2006)	.98	.97	.91	.98	.89	.79	-	-	-	.96	.92	.94	.99	.96	.87
Wilde (2006)	.95	.91	.85	.96	.85	.69	.95	.94	.93	-	-	-	.96	.91	.84
Vogt	.98	.96	.94	.98	.92	.79	.98	.97	.85	.95	.91	.84	-	-	-

Appendix C

RBANS subtest raw scores means and standard deviations

Subtest	Range	Carlozzi et al. (2008) *N* = 175	Duff et al. (2006) *N* = 796	Wilde (2006) *N* = 210	Vogt *N* = 345
List Learning	0 - 40	16.05 (5.49)	24.7 (5.9)	20.31 (5.82)	16.28 (5.48)
Story Memory	0 - 24	10.56 (5.03)	15.8 (4.5)	13.86 (4.30)	10.12 (5.04)
Figure Copy	0 - 20	13.24 (4.63)	18.2 (2.1)	14.33 (4.94)	17.60 (2.67)
Line Orientation	0 - 20	13.35 (5.18)	15.9 (3.6)	11.91 (4.79)	13.29 (4.59)
Picture Naming	0 - 10	9.06 (1.31)	9.56 (0.81)	8.87 (1.53)	9.10 (1.20)
Semantic Fluency	0 - 40	12.46 (5.29)	18.1 (4.7)	13.28 (5.04)	12.48 (4.87)
Digit Span	0 - 16	8.70 (2.41)	11.46 (2.79)	8.68 (2.42)	8.52 (2.15)
Coding	0 - 89	20.47 (13.45)	35.9 (10.7)	17.84 (11.67)	27.82 (11.92)
List Recall	0 - 10	1.28 (1.75)	5.0 (2.6)	2.78 (2.46)	1.32 (1.99)
List Recognition	0 – 20	16.13 (3.02)	18.9 (1.6)	17.40 (2.58)	15.77 (3.03)
Story Recall	0 - 12	4.02 (3.37)	7.9 (3.0)	6.02 (2.82)	3.63 (3.15)
Figure Recall	0 - 20	6.01 (4.92)	12.9 (4.3)	8.82 (4.93)	5.20 (5.55)

Note: Duff et al., 2010 reported subtest scores as standard scores not raw scores in publication, thus, omitted from table.