

# Bi-factor analyses of the Brief Test of Adult Cognition by Telephone

Brandon E. Gavett<sup>a,\*</sup>, Paul K. Crane<sup>b</sup> and Kristen Dams-O'Connor<sup>c</sup>

<sup>a</sup>*Department of Psychology, University of Colorado at Colorado Springs, Colorado Springs, CO, USA*

<sup>b</sup>*Department of Medicine, University of Washington, Seattle, WA, USA*

<sup>c</sup>*Department of Rehabilitation Medicine, Mount Sinai School of Medicine, New York, NY, USA*

## Abstract.

**BACKGROUND:** Telephone cognitive batteries are useful for large-scale screening and epidemiological studies, but their brevity and lack of content depth may cause psychometric limitations that hinder their utility.

**OBJECTIVE:** The current study addressed some of these limitations by rescaling the Brief Test of Adult Cognition by Telephone (BTACT; Tun & Lachman, 2006) using modern psychometric methods.

**METHODS:** Archival data were obtained from a national sample of 4,212 28 to 84-year-old volunteers in the National Survey of Midlife Development in the United States (Ryff et al., 2007) Cognitive Project (Ryff & Lachman, 2007). We fit a bi-factor model to a combination of item-level, subscale-level, and scale-level data.

**RESULTS:** The best fitting model contained a general factor and secondary factors capturing test-specific method effects or residual correlations for Number Series, Red/Green Test, and the Rey Auditory Verbal Learning Test. Factor scores generated from this model were compared with conventional BTACT scores. Important score differences (i.e., >0.3 standard deviation units) were found in 28% of the sample. The bi-factor scores demonstrated slightly superior validity than conventional BTACT scores when judged against a number of clinical and demographic criterion variables.

**CONCLUSIONS:** Modern psychometric approaches to scoring the BTACT have the benefit of linear scaling and a modest criterion validity advantage.

Keywords: Telephone, cognitive assessment, modern psychometrics, bi-factor model, neuropsychology

## 1. Introduction

Neuropsychological assessment involves the formal measurement of various domains of cognitive functioning using standardized tests, typically through face-to-face test administration. Although in-person assessment is often necessary or desirable for clinical, forensic, and research purposes, it may be undesirable or unfeasible in some situations, such as large-scale research studies. Telephone assessment of cognitive functioning provides an alternative method for

evaluating cognition in diverse populations and addresses many of the circumstances in which in-person evaluations are impractical or impossible. For instance, telephone assessment may help increase sample size and diversity by recruiting participants from a larger geographic region than would be feasible for in-person evaluations. Advances in telecommunications, including the widespread availability of interactive voice response systems, voice recognition and recording technologies, and computer-assisted telephone interviewing, may make telephone survey research even more advantageous than it has been in the past by improving the quality of data collected (Kempf & Remington, 2007). Telephone-assisted data collection allows for greater data quality control compared to mail

\*Address for correspondence: Brandon E. Gavett, Department of Psychology, 1420 Austin Bluffs Parkway, Colorado Springs, CO 80918, USA. Tel.: +1 719 255 4135; Fax: +1 719 255 4166; E-mail: bgavett@uccs.edu.

survey studies and is more cost-effective and efficient than in-person assessments (Lavrakas, 1993). Telephone cognitive instruments have been developed to serve as brief screening tools for dementia and other cognitive difficulties (e.g., Brandt et al., 1993; Crooks, Clark, Petitti, Chui, & Chiu, 2005; Gallo & Breitner, 1995). Some of the more commonly used telephone batteries include the Telephone Interview of Cognitive Status (Brandt et al., 1993) and the Brief Test of Adult Cognition by Telephone (BTACT; Tun & Lachman, 2006; Lachman & Tun, 2008). For a more comprehensive discussion, see Lachman and Tun (2008).

The benefits of telephone cognitive evaluations may be most apparent in epidemiological studies and other large-scale survey studies (Wilson & Bennett, 2005; Wolfson et al., 2009). In particular, a study that requires participants to undergo in-person assessment at the study site necessarily limits its sample to those individuals who have the health, mobility, and resources to travel. More generally, unless a study that utilizes in-person evaluations can provide transportation to all participants or conduct home visits, random sampling may be compromised as a result of sampling bias. The consequence of this sampling bias may be the under-sampling or exclusion of populations that are more impaired than those who are enrolled in the study (e.g., Dura & Kiecolt-Glaser, 1990). Individuals who do not participate in health studies and those who are lost to attrition tend to have more impaired functioning, poorer self-rated health, poorer cognitive functioning, and are more likely to live in rural areas (e.g., Matthews et al., 2004; Nummela et al., 2011). In large-scale studies of health and cognition, selection bias may be minimized through the use of remote assessment methods.

Large-scale research projects are not the only situation in which telephone cognitive batteries are advantageous. Telephone assessment has been used to track rehabilitation outcomes after discharge (Guerini et al., 2008; Jones, Miller, & Petrella, 2002; Lysack, Neufeld, Mast, MacNeill, & Lichtenberg, 2003; Worthington, Matthews, Melia, & Oddy, 2006), including cognitive outcomes (Dombrov, Drew-Cates, & Serdars, 1998). In the current era of health care reimbursement, hospitals face pressures to discharge patients from inpatient rehabilitation units before complete recovery can be achieved (Gillen, Tennen, & McKee, 2007). In order to track continued recovery post-discharge, outpatient visits may be recommended. However, in cases where continued monitoring of cognitive recovery is warranted, telephone-based follow-up may be a less costly and more efficient alternative to outpatient visits.

Given the large and rapidly growing population of older adults, there are many benefits to having a psychometrically sound cognitive assessment instrument that can be administered by telephone. However, limitations of telephone cognitive assessment can also pose a challenge to the quality of the data collected, both in terms of psychometrics and in terms of content. Visual stimuli cannot be presented over the phone, so telephone batteries must rely solely on tests that can be completed verbally, thus limiting the ability of the test to capture all of the constructs typically measured during an in-person neuropsychological assessment (Crooks, Petitti, Robins, & Buckwalter, 2006). Many telephone test batteries were designed to be administered in a brief period of time (30 minutes or less), which can make it difficult to include a sufficient number of quality items to sample a wide range of abilities and to measure ability with acceptable reliability. Accordingly, many existing telephone cognitive assessment batteries were designed to identify the absence or presence of specific deficits (e.g., Alzheimer's disease and other neurodegenerative conditions) and may be suitable only as screening tools for milder degrees of cognitive impairment (Duff, Beglinger, & Adams, 2009). However, because they have not been designed to measure cognitive functioning more broadly in the general population, most existing telephone cognitive tests cannot be expected to detect subtle differences in cognitive ability among individuals with average or better abilities (e.g., Knopman et al., 2010).

Several telephone cognitive batteries have been developed with a goal of providing a more comprehensive measure of cognition across the lifespan. For instance, the BTACT was designed to avoid some of the psychometric and methodological limitations of other telephone-administered cognitive assessments, such as floor and ceiling effects. The BTACT was designed for use in the National Survey of Midlife Development in the United States (MIDUS II; Brim, Ryff, & Kessler, 2004) Cognitive Project (Ryff & Lachman, 2007). Although it is not a lengthy test battery (estimated administration time of 15 minutes), its items were chosen to encompass a wider range of ability levels than batteries developed specifically for dementia screening (Tun & Lachman, 2006). As such, the BTACT may be a promising tool for use in large-scale research studies and projects that seek to measure a wider range of cognitive ability levels. The BTACT has been recommended for inclusion in the National Institutes of Health Common Data Elements for traumatic brain injury (TBI) and is currently being piloted by the National Institute

of Disability and Rehabilitation Research funded TBI Model Systems for potential inclusion in this prospective longitudinal study of TBI outcomes.

Traditionally, the BTACT is scored using classical test theory, in which correct responses are totaled to arrive at a final score. This approach makes important untested assumptions about relationships between each of the items and the overall construct measured by the test. Total scores are often used in statistical models that assume linear scaling properties, such that a unit difference towards the top end of the scale has the same implication as a unit difference in the middle of the scale or towards the bottom of the scale. Although the tradition has been to score cognitive tests this way, this may not be the optimal approach to scoring, as the above assumptions are often violated (see for example Crane et al., 2008). When tracking changes in cognition over time, linear scaling properties are highly desirable (Mungas & Reed, 2000). It would therefore be advantageous to develop a method for scoring the BTACT in such a way that it possesses linear scaling properties.

One such approach to ensuring linear scaling properties is through item response theory (IRT) scoring (Hambleton & Swaminathan, 1985; Hambleton, Swaminathan, & Rogers, 1991). IRT is widely used in educational testing and an introductory text specifically targeted at psychologists is available (Embretson & Reise, 2000). However, IRT models also make important assumptions that may be violated. A crucial assumption of such models is the assumption of local independence; that is, that the correlations between test items can be appropriately modeled as being due to a single underlying factor. After extracting the single common factor, there should be minimal residual correlations between pairs of items. This assumption may be violated in the face of important method effects, such as having the same list of words used for encoding and recall tasks in a neuropsychological battery (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

When the unidimensional IRT model is not consistent with the data (e.g., if method effects are present), a more complicated model is necessary. In this case, the bi-factor model may be especially useful (Gibbons & Hedeker 1992; Gibbons et al. 2007; McDonald, 1999; Reise, Morizot, & Hays, 2007). The bi-factor model includes a general factor that reflects the ability assessed by all of the items in the test, as well as secondary factors that reflect substantive sub-domains or method effects. The general factor in bi-factor models reflects the same construct intended to be summarized by the total score from classical test theory; that is, a summary of the

overall ability level measured by the test. Unlike the classical test theory score, however, scores from the general factor in bi-factor models have linear scaling properties, making them desirable for use in many applications.

The goal of the current study was to identify appropriate factor structures for the BTACT to determine an appropriate method for applying modern psychometric approaches. Because the BTACT includes subtests that use a common method to evaluate specific abilities, we hypothesized that the local independence assumption of IRT would be violated, thus requiring a bi-factor structure to best model the data (DeMars, 2006). We also hypothesized that the factor scores derived from best fitting model would, in addition to having linear scaling properties, have superior criterion validity compared to the conventional approach to scoring the BTACT.

## 2. Method

### 2.1. Participants

Participants were volunteers in the MIDUS-II Cognitive Project (Ryff & Lachman, 2007), a follow-up to a national survey of non-institutionalized adults selected by random digit dialing (Brim, Ryff, & Kessler, 2004). Part of the MIDUS-II Cognitive Project, completed between the years of 2004 and 2006, included a computer-assisted telephone cognitive assessment using the BTACT in a sample of 4,212 participants (85% of the MIDUS-II sample). The MIDUS-II Cognitive Project data are publicly available through the Inter-University Consortium for Political and Social Research website and were used for the current study.

Of the 4,212 participants in the MIDUS-II Cognitive Project, 234 (5.6%) were excluded from the current study due to questionable test validity on any portion of the cognitive evaluation. The MIDUS-II researchers made the determination of questionable test validity and because we were unable to determine the specific reasons for questionable validity, we chose to exclude all cases that were flagged for validity concerns. The remaining 3,978 participants ranged in age from 28 to 84 years ( $M = 55.8$ ,  $S. D. = 12.3$ ) and included 2,133 (54%) women and 1,840 (46%) men; sex was not available for 5 participants.

### 2.2. Measures

All measures are components of the BTACT and are described below.

*Rey Auditory Verbal Learning Test (RAVLT)* (Rey, 1964, Taylor, 1959). The RAVLT paradigm used in the BTACT consists of one 15-item immediate recall trial and a delayed recall trial. After checking for adequate hearing, the 15-item word list is read to the participant with a one-second pause in between each word. The participants are then given 90 seconds to freely recall as many words as possible from the list, in any order. Approximately 15 minutes later, participants are asked to freely recall as many words as possible from the list without any cues from the examiner. Correct responses, intrusions, and repetitions are recorded.

*Digits Backward*. In this task, participants are read a series of digits, beginning with a length of two, and are asked to orally reproduce the digits in reverse order. If participants incorrectly respond to a trial, they are given a second trial of the same span length. If both trials at the same span length are failed, the test is discontinued; otherwise, the span length increases up to a maximum of 8 digits. One point is earned for each span length that is successfully completed; possible raw scores range from 0 to 7.

*Category Fluency (Animals)*. Participants are asked to verbally state as many animals as possible in 60 seconds. One point is awarded for each unique item that belongs to the animal category.

*Red/Green Test*. The Red/Green Test has three parts: baseline normal, baseline reverse, and mixed switching. During the baseline normal trial, participants hear either the word “red” or the word “green,” and are asked to respond with either “stop” or “go,” respectively. During the baseline reverse phase, participants are asked to respond by saying “go” to a “red” stimulus and “stop” to a “green” stimulus. This is followed by a mixed phase, where the task demands alternate unpredictably between the normal (18 trials) and reverse (14 trials) conditions. Responses are scored as correct or incorrect and response latencies are measured in milliseconds (see Tun & Lachman, 2008 for more detail). Because reaction time (RT) was measured over the telephone, which may introduce important confounds for the accurate measurement of RT, we relied solely on accuracy data for analysis of Red/Green Test performance.

*Number Series*. For this test, participants are read a sequence of five numbers and asked to provide the sixth number in the series. Successful performance on this test requires the ability to recognize the pattern in the five number stimuli and use that pattern to derive the sixth number. There are five different series presented and one point is awarded for each correct response. Each

of the five Number Series items was analyzed separately as a dichotomous variable, where 0 = incorrect and 1 = correct.

*Backward Counting*. This task requires participants to count backwards by one, starting at 100, as quickly as possible, for 30 seconds. One point is awarded for each correctly sequenced number; the total score is 100 minus the sum of the last number reached plus the number of errors.

### 2.3. Data analysis

Data analysis consisted of two phases: a model building phase and a model validation phase. In the model building phase, we examined the fit of several hypothesized models to the BTACT data. In the model validation phase, we scored the BTACT using the best fitting model and compared these scores to those produced by conventional BTACT scoring to evaluate criterion validity.

Prior to analysis, we re-coded item responses for each measure to generate categorical variables with a maximum of 10 categories, as that is the maximum number that Mplus can handle. Specific re-coding schemata applied are described below and shown in Appendix A.

For both the immediate and delayed recall RAVLT trials, we created three variables to capture primacy, middle, and recency effects, given that recall based on word position in a list may capture different but overlapping memory abilities (Gavett & Horwitz, 2012). The total number of words from the first five list positions was summed to create a “primacy” score, the second five for a “middle” score, and the final five for a “recency” score. Between the immediate and delayed trials, there were a total of six RAVLT variables used in the analyses.

We used scores from the mixed phase of the Red/Green Test to create four variables based on the number of correct responses to the normal and reverse trials. We separately analyzed frequency of correct responses from the “switch” conditions (i.e., any trial that required a response type [normal or reverse] that was discrepant from the previous trial) and the “normal” conditions (i.e., any trial that required a response of the same type [normal or reverse] that was required in the previous trial) for both Normal and Reverse trials. This led to the generation of the following variables: Normal/Switch (items 9, 19, and 29), Reverse/Switch (items 4, 15, and 24), Normal/Other (items 1–3, 10–14, 20–23, and 30–32), and Reverse/Other (items 5–8, 16–18, 25–28). The two “Switch” variables ranged

from 0–3 correct. The two “Other” variables were recoded as detailed in Appendix A.

We recoded the Category Fluency and Backward Counting scores into ordinal variables based on deciles, according to Appendix A.

### 2.3.1. Model building

All confirmatory factor analysis (CFA) models were tested using a robust weighted least squares estimator (WLSMV) in Mplus version 6.11 (Muthén & Muthén, 2010) and model fit was judged using the Chi square test of model fit, comparative fit index (CFI), Tucker Lewis index (TLI) and root mean square error of approximation (RMSEA). With our sample size and associated statistical power, the Chi square test will almost certainly be statistically significant (suggesting model misfit), but is included for reference. Higher CFI and TLI values, especially those  $\geq 0.95$ , are suggestive of good fit (Bentler, 1990; Hu & Bentler, 1999). Lower RMSEA values, typically below 0.06, are suggestive of good model fit (Hu & Bentler, 1999).

We used a variety of methods to address the question of whether a single factor model was consistent with the data. We first fit a single factor CFA model; fit statistics suggested important model misspecification (see Results). We also evaluated Eigenvalues from exploratory factor analysis; the resulting scree plot is shown in Appendix B. Although there was a prominent first Eigenvalue, several additional Eigenvalues were larger than would be anticipated if the scale was sufficiently unidimensional. The ratio of the first (5.15) to the second (2.12) Eigenvalue was only 2.4, well less than the 4.0 rule of thumb consistent with a single factor structure (Reeve et al., 2007). Several Eigenvalues were greater than 1. All of these considerations suggested the need to evaluate more complicated structures than a single factor model (Lai, Crane, & Cella, 2006).

Our initial bi-factor model candidate considered secondary domains for the five Number Series items, the four Red/Green items, and the six RAVLT items. Fit for this model was improved compared with the single factor model (see Results). Factor loadings made sense except for a negative loading for one of the word list items on the word list domain.

Our second and final bi-factor model candidate was guided by modification indices for the single factor CFA model. The three largest modification indices were for the RAVLT immediate and delayed recall primacy, middle, and recency scores. These findings, along with the negative loading from our initial bi-factor model, prompted us to specify theoretically justifiable relation-

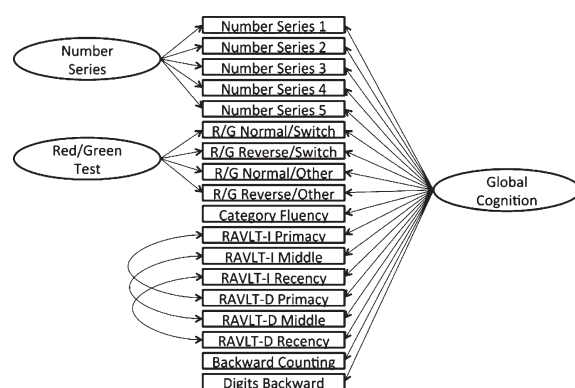


Fig. 1. The final bi-factor model found to provide the best fit to the BTACT data. R/G = Red/Green Test; RAVLT = Rey Auditory Verbal Learning Test; I = Immediate Recall; D = Delayed Recall.

ships among the six RAVLT indicators as three residual correlations rather than a single underlying factor. We show the final factor structure in Fig. 1. This model had excellent fit statistics (see Results). We extracted factor scores from the final bi-factor model for use in subsequent analyses.

### 2.3.2. Model validation

Although the conventional scoring of the BTACT derives separate factors for episodic memory and executive functioning, our modeling focused on development of a single global score rather than separate scores for episodic memory and executive functioning. We therefore focused on comparisons between global composite scores. R version 2.15.1 (R Core Team, 2012) was used to perform model validation analyses.

The conventional scoring of the BTACT uses  $z$ -score averaging to generate a global cognitive ability composite score. This score is generated by first deriving a  $z$ -score for each of the following variables based on the mean and the standard deviation of the overall MIDUS-II Cognitive Project sample: RAVLT (sum of Immediate and Delayed), Digits Backward, Category Fluency, Number Series, and Backward Counting. These five  $z$ -scores are then averaged and re-standardized ( $M = 0$ ,  $SD = 1$ ) to produce a composite  $z$ -score for the BTACT. In order to ensure that both the bi-factor scores and the conventional BTACT scores are compared on the same metric, we re-standardized the bi-factor global composite score to ensure that all variables being compared had the same mean of 0 and standard deviation of 1.

First, we directly compared the two BTACT scoring methods. We visually examined the relationship between bi-factor and conventional global cognition

scores to determine whether the relationship between the two variables was sufficiently linear to apply tests of association that assume linearity (e.g., Pearson's  $r$  correlation coefficient). We also sought to determine whether the magnitude of score differences between the two scoring approaches was substantial enough to warrant the use of the more sophisticated and computationally complex bi-factor scoring approach. We considered differences greater than 0.3 standard deviation units as meaningful. The rationale for this criterion is that 0.3 is the typical stopping rule for computerized adaptive testing, which de facto represents a tolerable amount of measurement error. Therefore, we examined the frequency with which the two scoring methods produced scores that differed by greater than  $>0.3$ .

To perform criterion validity comparisons, we split our sample into two groups based on the absence or presence of neurologic disease or injury known to affect cognition. We used survey responses from the MIDUS-II (Ryff et al., 2007) data to identify participants who self-reported a history of stroke (both lifetime and within the preceding 12 months), other neurologic illness (multiple sclerosis, epilepsy, or other neurological disorder; both lifetime and within the preceding 12 months), and lifetime history of serious head injury. We dichotomized our sample on the basis of these variables and used group means and standard deviations for each scoring method to compute the standardized difference (Cohen's  $d$ ; Cohen, 1988) between groups. Larger Cohen's  $d$  values indicate that there is a more substantial difference detected by the measure between groups with and without conditions known to cause cognitive impairment. We hypothesized that group differences would be more pronounced when global cognition was measured using bi-factor scoring compared to conventional scoring, which would indicate that the bi-factor scoring provides a more valid indicator of cognitive impairment.

We also obtained self-report data about perceived memory ability relative to age and relative to change in memory ability over the previous five years. Self-ratings of memory for age were coded as (1) "Excellent" (2) "Good" (3) "Average" (4) "Fair" and (5) "Poor." Self-reported memory change was rated as (1) "Improved a lot" (2) "Improved a little" (3) "Stayed the same" (4) "Gotten a little worse" and (5) "Gotten a lot worse." In addition to the ordinal scaling for self-reported memory, another criterion variable, educational attainment, was ordinally scaled using 12 levels, with a score of 1 representing 0–6 years of education and a score of 12 representing a Ph.D., MD, or

related terminal degrees. Because these variables were coded ordinally, we used a non-parametric measure of association, Kendall's tau (Kendall, 1938), to evaluate the criterion validity of the two scoring methods. We hypothesized that the bi-factor model would produce scores that were more strongly associated with self-reported memory and educational attainment than the conventional BTACT scores.

Finally, we evaluated associations with age due to the well-documented association between age and cognitive functioning (e.g., Hedden & Gabrieli, 2004). We treated age as a continuous variable for this analysis. We hypothesized that the bi-factor scoring method would demonstrate a stronger association with age than the conventional scoring method, as measured by Pearson's  $r$ .

### 3. Results

The fit of the single factor model was poor;  $\chi^2$  (135,  $N = 3,978$ ) = 15,016, baseline  $\chi^2$  (153,  $N = 3,978$ ) = 54,829, CFI = 0.73, TLI = 0.69; RMSEA = 0.17 (90% CI = 0.16–0.17). The first bi-factor model tested yielded improved, yet sub-standard fit indices;  $\chi^2$  (120,  $N = 3,978$ ) = 6, 824, CFI = 0.88, TLI = 0.84, RMSEA = 0.12 (90% CI = 0.12–0.12). The final bi-factor model fit the data well;  $\chi^2$  (123,  $N = 3,978$ ) = 1,697, CFI = 0.97, TLI = 0.96, RMSEA = 0.057 (90% CI = 0.054–0.059). We examined loadings on the general factor from both the single factor and bi-factor models, as shown in Table 1. Consistent with the discussion of the single factor model fit statistics above, we found dramatic differences – as much as 51% – in factor loadings between the single factor and the bi-factor models. In particular, the difference in loadings for immediate and delayed recall of the same words was impressive. Loadings on secondary domains and the item category thresholds are shown in Appendix C.

The relationship between the two global composite  $z$ -scores was strongly linear,  $r = 0.96$  (95% CI = 0.96–0.96), as can be seen in Fig. 2. We examined the frequency of meaningful differences (defined as  $>|0.3|$ ) between the conventional scores and the bi-factor scores. There were 627 (16%) participants with bi-factor  $z$ -scores that were  $>0.3$  standard deviation units higher than conventional  $z$ -scores and 462 (12%) participants with conventional  $z$ -scores that were  $>0.3$  standard deviation units higher than bi-factor  $z$ -scores, which means that 28% of the sample had score differences of greater than 0.3 standard deviation units. Examination of larger differences revealed that 7.6% of

Table 1  
Loadings on the general factor for the single-factor and bi-factor models

Indicator	Single factor model	Bi-factor model	Absolute difference	Percent difference %
Number Series 1	0.51	0.54	-0.03	-7
Number Series 2	0.47	0.56	-0.09	-19
Number Series 3	0.41	0.42	-0.01	-2
Number Series 4	0.53	0.57	-0.04	-8
Number Series 5	0.37	0.42	-0.06	-16
Red/Green Normal/Switch	0.33	0.27	0.06	19
Red/Green Reverse/Switch	0.32	0.34	-0.02	-5
Red/Green Normal/Other	0.33	0.31	0.02	6
Red/Green Reverse/Other	0.37	0.40	-0.03	-7
Category Fluency	0.42	0.56	-0.13	-32
RAVLT-I Primacy	0.72	0.46	0.26	36
RAVLT-I Middle	0.85	0.42	0.43	51
RAVLT-I Recency	0.36	0.26	0.10	27
RAVLT-D Primacy	0.76	0.52	0.24	32
RAVLT-D Middle	0.90	0.49	0.40	45
RAVLT-D Recency	0.47	0.40	0.06	14
Backward Counting	0.49	0.65	-0.17	-34
Digits Backward	0.43	0.54	-0.11	-27

Note. RAVLT = Rey Auditory Verbal Learning Test; I = Immediate; D = Delayed.

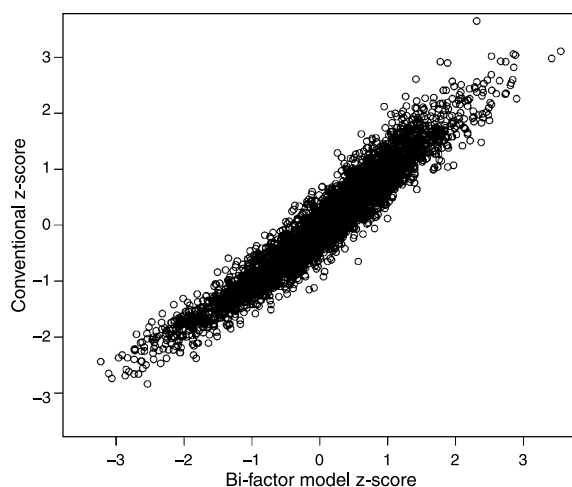


Fig. 2. Scatterplot of BTACT global composite scores produced by bi-factor scoring (x-axis) and conventional scoring (y-axis).

the sample differed by greater than 0.5 standard deviation units and 0.3% of the sample differed by greater than 1 standard deviation unit. See Fig. 3 for a graphical depiction of these differences.

The clinical significance of these score differences can be judged based on the data presented in Tables 2 and 3. For four of the five clinical groupings (lifetime history of stroke, 12-month history of stroke, 12-month history of other neurologic illness, and lifetime history of serious head injury), bi-factor scoring yielded somewhat larger group differences than did conventional scoring. This pattern was reversed for groupings based on lifetime history of other neurologic illness,

in that the conventional scoring produced somewhat larger group differences than the bi-factor scoring. In contrast to the group comparisons, the conventional scoring was more strongly correlated with a majority of the non-clinical variables (education, self-report of memory ability relative to age, self-report of memory change over five years). The bi-factor model produced scores that were more strongly related to age than the conventional model. Across all comparisons, the differences between bi-factor scoring and conventional scoring were modest.

#### 4. Discussion

The BTACT is a telephone-administered cognitive assessment instrument made up of six individual tests that measure immediate and delayed recall, working memory, semantic fluency, problem solving, mental control, and mental flexibility. The conventional approach to deriving a global composite score for the BTACT involves standardizing the average z-score of five of these tests (RAVLT Immediate plus Delayed recall, Digits Backward, Category Fluency, Number Series, and Backward Counting), an approach that is not guided by a theoretically-based latent structure for the abilities measured by the test. The purpose of this study was to apply modern psychometric techniques to the BTACT to determine whether consideration of the test's underlying factor structure can lead to improvements in the scaling of scores and test validity. We used a total of 18 variables, derived from the five tests used in

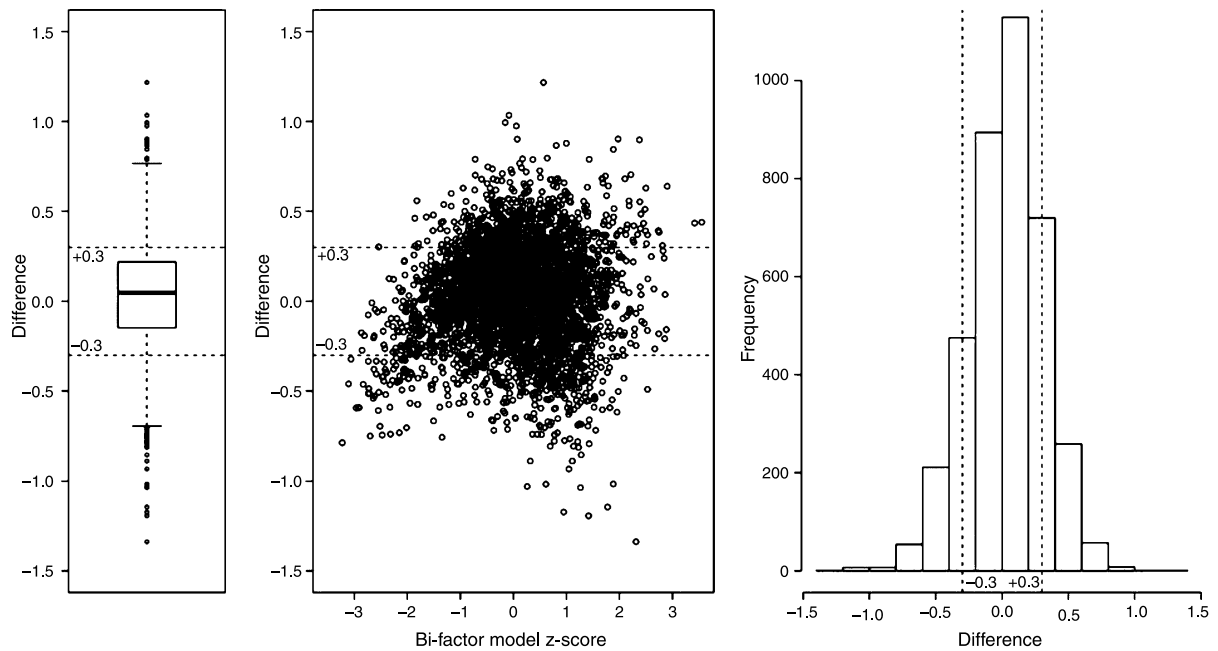


Fig. 3. Distribution of differences between bi-factor and conventional global composite scores on the BTACT, illustrated with a box-and-whisker plot (left panel), scatterplot (middle panel), and histogram (right panel). Demarcation points at +0.3 and -0.3 standard deviation units are provided to illustrate the proportion of test scores that differed by a meaningful amount.

Table 2  
Criterion group differences using conventional and bi-factor scoring

	Conventional scoring		Bi-factor scoring	
	<i>M</i> ( <i>SD</i> )	Cohen's <i>d</i> (95% CI)	<i>M</i> ( <i>SD</i> )	Cohen's <i>d</i> (95% CI)
Stroke – Lifetime				
Yes ( <i>n</i> = 117)	-0.74 (0.81)		-0.82 (0.92)	
No ( <i>n</i> = 3849)	0.03 (0.99)	0.78 (0.66–0.91)	0.03 (0.99)	<b>0.84 (0.71–1.00)</b>
Stroke – Last 12 months				
Yes ( <i>n</i> = 38)	-0.98 (0.80)		-1.05 (0.97)	
No ( <i>n</i> = 3468)	0.03 (1.00)	1.01 (0.79–1.23)	0.02 (1.00)	<b>1.06 (0.79–1.33)</b>
Neurologic Illness – Lifetime				
Yes ( <i>n</i> = 207)	-0.15 (1.02)		-0.16 (1.03)	
No ( <i>n</i> = 3733)	0.03 (0.99)	<b>0.18 (0.06–0.30)</b>	0.01 (1.00)	0.17 (0.05–0.29)
Neurologic Illness – Last 12 Months				
Yes ( <i>n</i> = 79)	-0.27 (1.01)		-0.31 (1.02)	
No ( <i>n</i> = 3427)	0.02 (1.00)	0.30 (0.10–0.48)	0.01 (1.00)	<b>0.32 (0.13–0.51)</b>
Serious Head Injury – Lifetime				
Yes ( <i>n</i> = 119)	0.00 (0.99)		-0.07 (1.00)	
No ( <i>n</i> = 3849)	0.01 (1.00)	0.01 (-0.15–0.17)	0.00 (1.00)	<b>0.07 (-0.08–0.22)</b>

the conventional scoring of the BTACT, along with the Red/Green Test. We found that a single-factor (unidimensional) model provides a poor fit for the BTACT data and derived a bi-factor model that fit the data well. In addition to a general factor (conceptualized as global cognitive functioning), secondary factors for the Number Series and Red/Green Tests and residual correlations between primacy, middle, and recency effects on the RAVLT were incorporated in the best fitting model.

Comparisons between BTACT scores that are produced by conventional scoring and those that are produced by the bi-factor model revealed that important score differences occurred in 28% of our sample despite a strong positive correlation between the two approaches. These differences occurred frequently enough to warrant more detailed validity comparisons. The results of these validity comparisons suggest that the bi-factor model produces scores that are some-



Table 3  
Correlations of conventional and bi-factor scoring with age, education, and self-reported memory

	Conventional scoring	Bi-factor scoring
Age ( <i>r</i> , 95% CI)	-0.43 (-0.45 to -0.41)	<b>-0.45 (-0.48 to -0.43)</b>
Education ( $\tau$ )	<b>0.31</b>	0.29
Self-reported memory for age ( $\tau$ )	<b>-0.15</b>	-0.14
Self-reported 5-year memory change ( $\tau$ )	<b>-0.02</b>	-0.01

what better able to separate those with and without a lifetime history of stroke, stroke within the last 12 months, other neurologic illness within the last 12 months, and a lifetime history of serious head injury. In addition, the bi-factor model produces scores that are more strongly correlated with age than the conventional scoring method. In contrast, the conventional approach better separated individuals with and without a lifetime history of neurologic illness and demonstrated a stronger correlation with education and self-ratings of memory ability relative to peers and changes in memory over five years. These results provide modest evidence to support our hypothesis that a bi-factor scaling of the BTACT can improve the measurement of global cognition relative to the atheoretical standard *z*-score approach to scaling. Although the conventional approach to scoring did show stronger associations with some of our criterion variables, these variables (lifetime history of other neurologic illness, education, and self-ratings of memory) may be considered less clinically relevant than those criterion variables that were more strongly associated with bi-factor scores (12-month history of stroke and other neurologic illness, lifetime history of stroke, lifetime history of serious head injury, age). For instance, self-reported memory has previously been shown to correlate poorly with objective measurements of memory (Schmidt, Berg, & Deelman, 2001). It should be noted that for all validity comparisons, differences between scoring methods were quite modest.

One important difference between the conventional approach to scoring the BTACT and the bi-factor model validated in this study is the inclusion of the Red/Green Test scores in the bi-factor model. The conventional approach to scoring the BTACT does not account for performance on this test. Therefore, another advantage of the bi-factor model is that the inclusion of this additional test, which appears to require set shifting, mental flexibility, response inhibition, and sustained

auditory attention, may add to the comprehensiveness of the global composite score produced by the bi-factor model. Relatively speaking, of course, inclusion of the Red/Green Test data in the composite score somewhat reduces the influence of memory on the total score. This may explain the attenuated strength of association with self-reported memory performance compared with standard scoring.

Although the criterion validity advantage of the scores produced by the bi-factor model over the conventional approach to scoring the BTACT was modest, there is nevertheless a strong psychometric advantage for using the bi-factor scores to interpret test performance. The factor scores generated by the bi-factor model possess linear measurement properties, which means that there is a direct relationship between differences in cognitive ability and differences in test scores (Mungas & Reed, 2000). Without linear scaling, a test primarily made up of easy items would be less able to differentiate between two individuals with above-average, yet different, cognitive abilities compared to its ability to identify the same magnitude of difference in below-average individuals. Linear scaling properties are also desirable in longitudinal applications because the same numeric change score reflects a unit of change in true ability level regardless of initial ability level (Crane et al., 2008; Mungas et al., 2010). However, because the MIDUS-II Cognitive Project data are cross-sectional, further research that collects longitudinal BTACT data is required to report on the longitudinal measurement properties of the bi-factor BTACT.

The absence of longitudinal data is an important limitation of the current study. One of the advantages of telephone cognitive assessment is that it may be more convenient than in-person assessment for performing repeated evaluations to monitor rehabilitation outcomes or the progression of a neurodegenerative disease. Future studies should examine the longitudinal measurement properties of the bi-factor BTACT scores. Telephone cognitive data themselves may be somewhat less reliable than data obtained via in-person evaluations, due to the relative lack of control over the testing environment and communication challenges that may be amplified without face-to-face communication. In addition, telephone cognitive assessment typically involves audio recording of responses, which has been shown to reduce the validity of test results (Constantinou, Ashendorf, & McCaffrey, 2002). However, these are limitations of telephone cognitive assessment in general and not specific to the findings reported here. Another limitation of this study is the

reliance upon self-report for criterion validity studies. Because the MIDUS-II is a study of healthy aging, specific clinical samples were not recruited, and medical documentation of the presence and severity of stroke, neurologic illness, and brain injury history was unavailable. The bi-factor model presented here should be studied in clinical samples with medically verified neurologic disease. In addition, no cognitive tests other than the BTACT were administered to participants in the MIDUS-II Cognitive Project, which prevented us from examining the convergent validity of the bi-factor BTACT scores. Nevertheless, the current results add to existing research on the BTACT and support its ability to provide a brief measure of global cognition across a wide range of ages and ability levels.

### Acknowledgments

The Mplus code used to generate bi-factor scores can be obtained by contacting BEG (E-mail: bgavett@uccs.edu.) or PKC (E-mail: pcrane@uw.edu.). Dr. Crane's time was supported by R01 AG 029672 (Crane). The authors would like to thank the Program Committee of the Advanced Psychometric Methods in Cognitive Aging Workshop (R13 AG030995; Mungas) for facilitating this collaboration.

### Declaration of interest

The authors have no conflicts of interest to declare.

### References

- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.
- Brandt, J., Welsh, K. A., Breitner, J. C., Folstein, M. F., Helms, M., & Christian, J. C. (1993). Hereditary influences on cognitive functioning in older men. A study of 4000 twin pairs. *Archives of Neurology*, *50*, 599-603.
- Brim, O., Ryff, C., & Kessler, R. (2004). *How healthy are we?: A national study of well-being at midlife*. Chicago: University of Chicago Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd Ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crane, P. K., Narasimhalu, K., Gibbons, L. E., Mungas, D. M., Haneuse, S., Larson, E. B., & van Belle, G. (2008). Item response theory facilitated cocalibrating cognitive tests and reduced bias in estimated rates of decline. *Journal of Clinical Epidemiology*, *61*, 1018-1027.
- Crooks, V. C., Clark, L., Petitti, D. B., Chui, H., & Chiu, V. (2005). Validation of multi-stage telephone-based identification of cognitive impairment and dementia. *BMC Neurology*, *5*, 8.
- Crooks, V. C., Petitti, D. B., Robins, S. B., & Buckwalter, J. G. (2006). Cognitive domains associated with performance on the telephone interview for cognitive status-modified. *American Journal of Alzheimer's Disease and Other Dementias*, *21*, 45-53.
- Constantinou, M., Ashendorf, L., & McCaffrey, R. J. (2002). When the third party observer of a neuropsychological evaluation is an audio-recorder. *The Clinical Neuropsychologist*, *16*, 407-412.
- DeMars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *Journal of Educational Measurement*, *43*, 145-168.
- Dombovy, M. L., Drew-Cates, J., & Serdars, R. (1998). Recovery and rehabilitation following subarachnoid hemorrhage: Part II long-term follow-up. *Brain Injury*, *12*, 887-894.
- Duff, K., Beglinger, L. J., & Adams, W. H. (2009). Validation of the modified telephone interview for cognitive status in amnesic mild cognitive impairment and intact elders. *Alzheimer Disease and Associated Disorders*, *23*, 38-43.
- Dura, J. R. & Kiecolt-Glaser, J. K. (1990). Sample bias in caregiving research. *Journal of Gerontology*, *45*, 200-204.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gallo, J. J. & Breitner, J. C. (1995). Alzheimer's disease in the NAS-NRC Registry of aging twin veterans, IV. Performance characteristics of a two-stage telephone screening procedure for Alzheimer's dementia. *Psychological Medicine*, *25*, 1211-1219.
- Gavett, B. E. & Horwitz, J. E. (2012). Immediate list recall as a measure of short-term episodic memory: Insights from the serial position effect and item response theory. *Archives of Clinical Neuropsychology*, *27*, 125-135.
- Gibbons, R. D. & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*, 423-436.
- Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., & Stover, A. (2007). Full-information item bi-factor analysis of graded response data. *Applied Psychological Measurement*, *31*, 4-19.
- Gillen, R., Tennen, H., & McKee, T. (2007). The impact of the inpatient rehabilitation facility prospective payment system on stroke program outcomes. *American Journal of Physical Medicine and Rehabilitation*, *86*, 356-363.
- Guerini, F., Frisoni, G. B., Marrè, A., Turco, R., Bellelli, G., & Trabucchi, M. (2008). Subcortical vascular lesions predict falls at 12 months in elderly patients discharged from a rehabilitation ward. *Archives of Physical Medicine and Rehabilitation*, *89*, 1522-1527.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory. Principles and applications*. Boston: Kluwer-Nijhoff.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hedden, T. & Gabrieli, J. D. (2004). Insights into the ageing mind: A view from cognitive neuroscience. *Nature Reviews Neuroscience*, *5*, 87-96.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.
- Jones, G. R., Miller, T. A., & Petrella, R. J. (2002). Evaluation of rehabilitation outcomes in older patients with hip fractures. *American Journal of Physical Medicine and Rehabilitation*, *81*, 489-497.
- Kempf, A. M. & Remington, P. L. (2007). New challenges for telephone survey research in the twenty-first century. *Annual Reviews of Public Health*, *28*, 113-126.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*, 81-93.

- Knopman, D. S., Roberts, R. O., Geda, Y. E., Pankratz, V. S., Christianson, T. J., Petersen, R. C., & Rocca, W. A. (2010). Validation of the telephone interview for cognitive status-modified in subjects with normal cognition, mild cognitive impairment, or dementia. *Neuroepidemiology*, *34*, 34-42.
- Lachman, M. E. & Tun, P. A. (2008). Cognitive testing in large scale surveys. Assessment by telephone. In S. Hofer & D. Alwin (Eds.). *Handbook of cognitive aging: Interdisciplinary perspectives* (pp. 506-522). Thousand Oaks, CA: Sage.
- Lai, J. S., Crane, P. K., & Cella, D. (2006). Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue. *Quality of Life Research*, *15*, 1179-1190.
- Lavrakas, P.J. (1993). *Telephone Survey Methods: Sampling, Selection, and Supervision*. Newbury Park, CA: Sage.
- Lysack, C. L., Neufeld, S., Mast, B. T., MacNeill, S. E., & Lichtenberg, P. A. (2003). After rehabilitation: An 18-month follow-up of elderly inner-city women. *American Journal of Occupational Therapy*, *57*, 298-306.
- Matthews, F. E., Chatfield, M., Freeman, C., McCracken, C., & Brayne, C. (2004). Attrition and bias in the MRC cognitive function and ageing study: An epidemiological investigation. *BMC Public Health*, *4*, 12.
- McDonald, R. P. (1999) *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mungas, D., Beckett, L., Harvey, D., Tomaszewski Farias, S., Reed, B., Carmichael, O., & DeCarli C. (2010). Heterogeneity of cognitive trajectories in diverse older persons. *Psychology and Aging*, *25*, 606-619.
- Mungas, D. & Reed, B. R. (2000). Application of item response theory for development of a global functioning measure of dementia with linear measurement properties. *Statistics in Medicine*, *19*, 1631-1644.
- Muthén, L.K. and Muthén, B.O. (1998-2010). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.
- Nummela, O., Sulander, T., Helakorpi, S., Haapola, I., Uutela, A., Heinonen, H., Valve, R., & Fogelholm, M. (2011). Register-based data indicated nonparticipation bias in a health study among aging people. *Journal of Clinical Epidemiology*, *64*, 1418-1425.
- Podskaoff, P. M., MacKenzie, S. B., Lee, J-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879-903.
- R Core Team. (2012). *R: A language and environment for statistical computing* [computer software]. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.R-project.org>.
- Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., & PROMIS Cooperative Group. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care*, *45*(Suppl 1), S22-31.
- Reise, S. P., Morizot, J. & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*(Suppl. 1), 19-31.
- Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.
- Ryff, C. Almeida, D. M. Ayanian, J. S. Carr, D. S., Cleary, P. D., Coe, C., & Williams, D. (2007). *National Survey of Midlife Development in the United States (MIDUS II), 2004-2006. ICPSR04652-v6*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Ryff, C. D. & Lachman, M. E. (2007). *National Survey of Midlife Development in the United States (MIDUS II): Cognitive Project, 2004-2006* [Data file]. ICPSR25281-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-07-13. doi:10.3886/ICPSR25281. Retrieved from <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/04652>.
- Schmidt, I. W., Berg, I. J., & Deelman, B. G. (2001). Relations between subjective evaluations of memory and objective memory performance. *Perceptual and Motor Skills*, *93*, 761-776.
- Taylor, E. M. (1959). *Psychological appraisal of children with cerebral deficits*. Cambridge, MA: Harvard University Press.
- Tun, P. A. & Lachman, M. E. (2006). Telephone assessment of cognitive function in adulthood: The Brief Test of Adult Cognition by Telephone. *Age and Ageing*, *35*, 629-632.
- Tun, P. A. & Lachman, M. E. (2008). Age differences in reaction time and attention in a national telephone sample of adults: Education, sex, and task complexity matter. *Developmental Psychology*, *44*, 1421-1429.
- Wilson, R. S. & Bennett, D. A. (2005). Assessment of cognitive decline in old age with brief tests amenable to telephone administration. *Neuroepidemiology*, *25*, 19-25.
- Wolfson, C., Kirkland, S. A., Raina, P. S., Uniat, J., Roberts, K., Bergman, H., & Meneok, K. (2009). Telephone-administered cognitive tests as tools for the identification of eligible study participants for population-based research in aging. *Canadian Journal of Aging*, *28*, 251-259.
- Worthington, A. D., Matthews, S., Melia, Y., & Oddy, M. (2006). Cost-benefits associated with social outcome from neurobehavioural rehabilitation. *Brain Injury*, *20*, 947-957.

**Appendix A**

Table A1  
Observed response frequencies and recoding of the “Normal/Other” trials of the Red/Green Test

Raw score	Frequency	Recoded value
0	3	0
1	1	0
3	3	0
4	2	0
5	2	0
6	7	0
7	8	0
8	12	1
9	10	2
10	27	3
11	64	4
12	104	5
13	104	6
14	378	7
15	3,253	8

Table A2  
Observed response frequencies and recoding of the “Reverse/Other” trials of the Red/Green Test

Raw score	Frequency	Recoded value
0	13	0
1	4	1
2	5	1
3	10	2
4	11	3
5	6	3
6	24	4
7	88	5
8	80	6
9	156	7
10	518	8
11	3,063	9

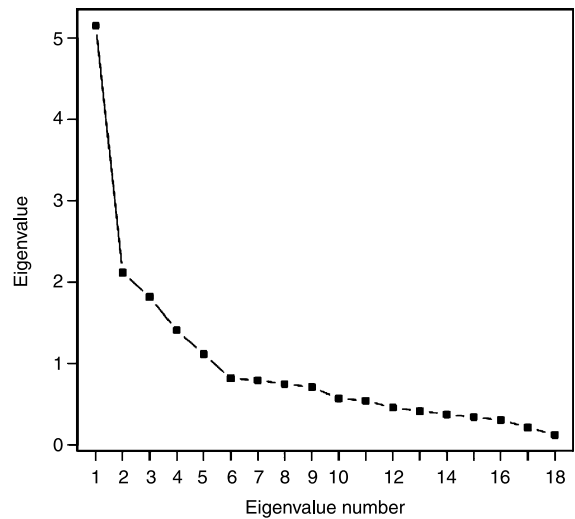
Table A3  
Recoding for Backwards Counting

Raw score	Recoded value
0–24	1
25–28	2
29–31	3
32–34	4
35–37	5
38–40	6
41–43	7
44–46	8
47–52	9
53–100	10

Table A4  
Recoding for Category Fluency

Raw score	Recoded value
0–11	1
12–14	2
15–16	3
17	4
18–19	5
20	6
21–22	7
23–24	8
25–27	9
28+	10

**Appendix B.** Scree plot of Eigenvalues from exploratory factor analysis.



**Appendix C.** Loadings and threshold values for final bi-factor model

	Estimate	S.E.	Est./S.E.	p-Value
Global	BY			
Number Series 1	0.54	0.019	28.171	0
Number Series 2	0.556	0.022	25.168	0
Number Series 3	0.422	0.022	19.473	0
Number Series 4	0.566	0.019	30.608	0
Number Series 5	0.424	0.021	19.861	0
Red/Green Reverse/Other	0.397	0.021	18.489	0
Red/Green Normal/Other	0.306	0.025	12.452	0
Red/Green Normal/Switch	0.265	0.033	8.006	0
Red/Green Reverse/Switch	0.337	0.025	13.697	0
Category Fluency	0.557	0.014	39.61	0
RAVLT-I Primacy	0.462	0.016	29.211	0
RAVLT-I Middle	0.422	0.016	26.448	0
RAVLT-I Recency	0.264	0.018	15.072	0
RAVLT-D Primacy	0.515	0.015	34.19	0
RAVLT-D Middle	0.493	0.015	32.427	0

	Estimate	S.E.	Est./S.E.	p-Value		Estimate	S.E.	Est./S.E.	p-Value
RAVLT-D Recency	0.403	0.016	24.722	0	Category Fluency \$2	-0.713	0.022	-32.659	0
Backward Counting	0.653	0.013	50.695	0	Category Fluency \$3	-0.356	0.02	-17.498	0
Digits Backward	0.538	0.015	36.054	0	Category Fluency \$4	-0.182	0.02	-9.128	0
Number Series	BY				Category Fluency \$5	0.15	0.02	7.513	0
Number Series 1	0.534	0.027	19.89	0	Category Fluency \$6	0.322	0.02	15.923	0
Number Series 2	0.267	0.033	8.023	0	Category Fluency \$7	0.63	0.021	29.463	0
Number Series 3	0.572	0.028	20.546	0	Category Fluency \$8	0.932	0.023	39.91	0
Number Series 4	0.563	0.027	21.054	0	Category Fluency \$9	1.377	0.028	48.336	0
Number Series 5	0.291	0.03	9.651	0	RAVLT-I Primacy\$1	-1.672	0.034	-48.998	0
Red/Green Test	BY				RAVLT-I Primacy\$2	-0.875	0.023	-38.21	0
Red/Green Reverse/Other	0.555	0.023	23.758	0	RAVLT-I Primacy\$3	-0.14	0.02	-7.038	0
Red/Green Normal/Other	0.644	0.023	27.51	0	RAVLT-I Primacy\$4	0.627	0.021	29.341	0
Red/Green Normal/Switch	0.722	0.029	24.825	0	RAVLT-I Primacy\$5	1.498	0.031	49.066	0
Red/Green Reverse/Switch	0.539	0.025	21.395	0	RAVLT-I Middle\$1	-1.102	0.025	-44.179	0
Global	WITH				RAVLT-I Middle\$2	-0.176	0.02	-8.811	0
Number Series	0	0	999	999	RAVLT-I Middle\$3	0.57	0.021	27.03	0
Red/Green Test	0	0	999	999	RAVLT-I Middle\$4	1.242	0.027	46.73	0
Number Series	WITH				RAVLT-I Middle\$5	1.901	0.04	47.071	0
Red/Green Test	0	0	999	999	RAVLT-I Recency\$1	-1.65	0.034	-49.079	0
RAVLT-I Primacy	WITH				RAVLT-I Recency\$2	-0.848	0.023	-37.344	0
RAVLT-D Primacy	0.543	0.014	39.823	0	RAVLT-I Recency\$3	-0.096	0.02	-4.819	0
RAVLT-I Middle	WITH				RAVLT-I Recency\$4	0.765	0.022	34.549	0
RAVLT-D Middle	0.659	0.013	50.053	0	RAVLT-I Recency\$5	1.72	0.035	48.759	0
RAVLT-I Recency	WITH				RAVLT-D Primacy\$1	-0.727	0.022	-33.202	0
RAVLT-D Recency	0.554	0.013	44.163	0	RAVLT-D Primacy\$2	-0.002	0.02	-0.095	0.924
Thresholds					RAVLT-D Primacy\$3	0.649	0.021	30.229	0
Number Series 1\$1	0.264	0.02	13.116	0	RAVLT-D Primacy\$4	1.303	0.027	47.557	0
Number Series 2\$1	-0.835	0.023	-36.88	0	RAVLT-D Primacy\$5	2.057	0.046	44.748	0
Number Series 3\$1	0.279	0.02	13.811	0	RAVLT-D Middle\$1	-0.618	0.021	-29.003	0
Number Series 4\$1	0.271	0.02	13.432	0	RAVLT-D Middle\$2	0.24	0.02	11.944	0
Number Series 5\$1	0.446	0.021	21.626	0	RAVLT-D Middle\$3	0.962	0.024	40.77	0
Red/Green Reverse/Other\$1	-2.72	0.092	-29.699	0	RAVLT-D Middle\$4	1.609	0.033	49.172	0
Red/Green Reverse/Other\$2	-2.541	0.074	-34.177	0	RAVLT-D Middle\$5	2.255	0.055	40.882	0
Red/Green Reverse/Other\$3	-2.407	0.064	-37.432	0	RAVLT-D Recency\$1	-0.583	0.021	-27.555	0
Red/Green Reverse/Other\$4	-2.247	0.055	-41.052	0	RAVLT-D Recency\$2	0.287	0.02	14.25	0
Red/Green Reverse/Other\$5	-2.089	0.047	-44.179	0	RAVLT-D Recency\$3	1.09	0.025	43.931	0
Red/Green Reverse/Other\$6	-1.745	0.036	-48.595	0	RAVLT-D Recency\$4	1.742	0.036	48.615	0
Red/Green Reverse/Other\$7	-1.55	0.032	-49.183	0	RAVLT-D Recency\$5	2.364	0.061	38.448	0
Red/Green Reverse/Other\$8	-1.283	0.027	-47.3	0	Backward Counting\$1	-1.205	0.026	-46.119	0
Red/Green Reverse/Other\$9	-0.739	0.022	-33.622	0	Backward Counting\$2	-0.823	0.023	-36.534	0
Red/Green Normal/Other\$1	-2.482	0.07	-35.628	0	Backward Counting\$3	-0.496	0.021	-23.861	0
Red/Green Normal/Other\$2	-2.343	0.06	-38.912	0	Backward Counting\$4	-0.193	0.02	-9.656	0
Red/Green Normal/Other\$3	-2.255	0.055	-40.882	0	Backward Counting\$5	0.06	0.02	2.998	0.003
Red/Green Normal/Other\$4	-2.078	0.047	-44.376	0	Backward Counting\$6	0.334	0.02	16.453	0
Red/Green Normal/Other\$5	-1.813	0.038	-48.042	0	Backward Counting\$7	0.61	0.021	28.667	0
Red/Green Normal/Other\$6	-1.546	0.031	-49.178	0	Backward Counting\$8	0.859	0.023	37.697	0
Red/Green Normal/Other\$7	-1.358	0.028	-48.158	0	Backward Counting\$9	1.296	0.027	47.446	0
Red/Green Normal/Other\$8	-0.907	0.023	-39.179	0	Digits Backward\$1	-2.746	0.095	-29.035	0
Red/Green Normal/Switch\$1	-2.839	0.106	-26.722	0	Digits Backward\$2	-1.966	0.043	-46.177	0
Red/Green Normal/Switch\$2	-2.353	0.061	-38.683	0	Digits Backward\$3	-1.132	0.025	-44.794	0
Red/Green Normal/Switch\$3	-1.449	0.03	-48.85	0	Digits Backward\$4	-0.251	0.02	-12.484	0
Red/Green Reverse/Switch\$1	-2.592	0.079	-32.911	0	Digits Backward\$5	0.439	0.021	21.335	0
Red/Green Reverse/Switch\$2	-1.916	0.041	-46.872	0	Digits Backward\$6	0.931	0.023	39.89	0
Red/Green Reverse/Switch\$3	-0.912	0.023	-39.32	0	Digits Backward\$7	1.422	0.029	48.674	0
Category Fluency \$1	-1.263	0.027	-47.03	0					