



Published in final edited form as:

*Rehabil Psychol.* 2017 November ; 62(4): 435–442. doi:10.1037/rep0000183.

## Factor Structure of the NIH Toolbox Cognition Battery in Individuals with Acquired Brain Injury

David S. Tulsky<sup>1,2</sup>, James A. Holdnack<sup>2</sup>, Matthew L. Cohen<sup>2,3</sup>, Robert K. Heaton<sup>4</sup>, Noelle E. Carlozzi<sup>5</sup>, Alex W. K. Wong<sup>6</sup>, Aaron J. Boulton<sup>2</sup>, and Allen W. Heinemann<sup>7,8</sup>

<sup>1</sup>Departments of Physical Therapy and Psychological and Brain Sciences, University of Delaware, Newark, DE

<sup>2</sup>The Center for Health Assessment Research and Translation, University of Delaware, Newark, DE

<sup>3</sup>Departments of Communication Science and Disorders and Psychological and Brain Sciences, University of Delaware, Newark, DE

<sup>4</sup>Department of Psychiatry, University of California - San Diego School of Medicine, San Diego, CA

<sup>5</sup>Center for Clinical Outcomes Development and Application and Department of Physical Medicine and Rehabilitation, University of Michigan Medical Center, Ann Arbor, MI

<sup>6</sup>Department of Occupational Therapy and Neurology, Washington University School of Medicine, St. Louis, MO

<sup>7</sup>Department of Physical Medicine and Rehabilitation; Feinberg School of Medicine, Northwestern University, Chicago, IL

<sup>8</sup>Center for Rehabilitation Outcomes Research, Shirley Ryan AbilityLab (formerly the Rehabilitation Institute of Chicago)

### Abstract

**Objective**—The National Institutes of Health Toolbox Cognition Battery (NIHTB-CB) measures reading, vocabulary, episodic memory, working memory, executive functioning and processing speed. While previous research has validated the factor structure in healthy adults, the factor structure has not been examined in adults with neurological impairments. Thus, this study evaluated the NIHTB-CB factor structure in individuals with acquired brain injury.

**Method**—A sample of 392 individuals (ages 18-84) with acquired brain injury ( $n=182$  TBI,  $n=210$  stroke) completed the NIHTB-CB along with neuropsychological tests as part of a larger, multi-site research project.

**Results**—Confirmatory factor analyses supported a 5-factor solution that included reading, vocabulary, episodic memory, working memory, and processing speed/executive functioning. This structure generally held in TBI and stroke subsamples as well as in subsamples of those with severe TBI and stroke injuries.

**Conclusions**—The factor structure of the NIHTB-CB is similar in adults with acquired brain injury to adults from the general population. We discuss the implications of these findings for clinical practice and clinical research.

### Keywords

Neuropsychology; Cognition; Factor Analysis; Brain Injuries; Stroke

The National Institutes of Health Toolbox of Neurological and Behavioral Functioning (NIHTB) is a standardized set of measures of cognition, emotion, motor, and sensory function. As a common data element, it was developed to provide a common measurement framework to facilitate the synthesis of results across studies (Gershon et al., 2010; Gershon, Wagster, et al., 2013). The NIHTB-Cognition Battery (NIHTB-CB) was designed for use with individuals 3 to 85 years of age and takes less than 30 minutes to administer. Developers obtained iterative feedback from research scientists and clinicians (Weintraub et al., 2013; Weintraub et al., 2014). Previous investigations have supported the validity of the NIHTB-CB measures of episodic memory (Bauer et al., 2013; Dikmen et al., 2014), working memory (Tulsky et al., 2014; Tulsky et al., 2013), reading (Gershon et al., 2014; Gershon, Slotkin, et al., 2013), vocabulary (Gershon et al., 2014; Gershon, Slotkin, et al., 2013), inhibitory control (Zelazo et al., 2013; Zelazo et al., 2014), cognitive flexibility (Zelazo et al., 2013; Zelazo et al., 2014), and processing speed (Carlozzi et al., 2014). The NIHTB-CB produces T-scores for each measure as well as three composite T-scores – crystallized cognition, fluid cognition, and overall cognition (Akshoomoff et al., 2013; Heaton et al., 2014). These scores can be adjusted for age, sex, education, and race/ethnicity (Casaletto et al., 2015).

Validity is not a property of a test itself, but rather how test scores are to be interpreted. For this reason, interpretation of scores must be validated for the populations with which they will be used. Test interpretation depends on relationships between manifest and latent variables, and exploratory and confirmatory factor analyses help elucidate these relationships. Confirmatory factor analysis (CFA) is useful when there are theoretical or empirical reasons to expect a factor structure, or one wants to evaluate competing factor models (Stevens, 1996). CFA is used to evaluate *a priori* hypotheses regarding the relationships between measured variables (test scores) and latent variables (domains of functioning), and is often used to support the interpretation of scores from a battery of cognitive tests (Benson, Hulac, & Kranzler, 2010; Bowden, Carstairs, & Shores, 1999; Holdnack, Xiaobin, Larrabee, Millis, & Salthouse, 2011; Tulsky & Price, 2003; Ward, Bergman, & Hebert, 2012; Weiss, Keith, Zhu, & Chen, 2013a, 2013b). “Goodness-of-fit” statistical criteria allow comparison of competing models (Tulsky & Price, 2003) and provide evidence of convergent and discriminant validity (Mungas et al., 2014; Mungas et al., 2013).

The factor structure of a cognitive test battery may vary because of demographic or clinical factors; thus, interpretation of battery scores requires knowledge of the factor structure in the population for which it was developed. Robust factor structures that are consistent across demographic and clinical groups are said to be “invariant”; CFA can help establish variance

or invariance across samples (Benson et al., 2010; Bowden, Lissner, McCarthy, Weiss, & Holdnack, 2007; Bowden, Saklofske, & Weiss, 2011; Mungas et al., 2014; Niileksela, Reynolds, & Kaufman, 2013; Taub, McGrew, & Witta, 2004; Tulsky & Price, 2003). For example, CFA has been used to evaluate the stability of factor structures in samples of individuals with traumatic brain injury (TBI)(van der Heijden & Donders, 2003), schizophrenia (Dickinson, Iannone, & Gold, 2002), substance abuse (Bowde et al., 2001), and heterogeneous clinical samples (Bodin, Pardini, Burns, & Stevens, 2009; Bowden, Weiss, Holdnack, Bardenhagen, & Cook, 2008; Weiss et al., 2013a, 2013b).

Two reports have described the factor structure of the NIHTB-CB. The first evaluated the convergent and discriminant validity of the cognitive tests in typically-developing children, ages 3-15 (Mungas et al., 2013). The authors analyzed the factor structure of the NIHTB-CB along with other, established neuropsychological tests that presumably measure the same constructs as the NIHTB-CB. The results supported a 5-factor model of cognitive function consisting of vocabulary, reading, episodic memory, working memory, and executive/processing speed factors for ages 8-15 years, and a 3-factor model for ages 3-7 consisting of vocabulary, reading, and fluid abilities (Mungas et al., 2013). These results support the construct validity of the NIHTB-CB in children without identified health problems. Each NIHTB-CB test measured the constructs they were intended to measure although the executive functioning tests were highly related to processing speed.

The second NIHTB-CB factor analysis study used a sample of healthy adults from the general population, 20-85 years of age. Participants completed a battery of neuropsychological tests that presumably measured the same constructs as the NIHTB-CB (Mungas et al., 2014). The results supported a 5-factor model that included reading, vocabulary, episodic memory, working memory, and processing speed/executive functioning (Mungas et al., 2014). A 6-factor solution that differentiated processing speed and executive functioning was not supported by fit statistics.

To our knowledge, no study has reported the factor structure of NIHTB-CB in clinical populations. This report is published with others that evaluate the use of the NIHTB-CB in individuals with traumatic brain injury (TBI), stroke, and spinal cord injury [MASKED]. Validation in populations with acquired brain injury, specifically TBI and stroke, is important because these are common neurological conditions that can result in cognitive impairment. This investigation has three aims: (1) to describe the latent variables underlying the NIHTB-CB by performing a CFA that includes the NIHTB-CB and more established neuropsychological tests, (2) to examine the factor structure and validate the test battery in adults with acquired brain injury, and (3) evaluate the effects of injury type and severity on factor structure.

## Method

### Participants

Individuals with acquired brain injury from TBI or stroke were recruited for a multi-site study that we described previously [MASKED]. They were recruited through registries at the Shirley Ryan AbilityLab (formerly the Rehabilitation Institute of Chicago), Washington

University School of Medicine, and the University of Michigan. We characterized the severity of the TBI [MASKED] or stroke [MASKED] by reviewing medical records.

Two hundred and ten participants sustained a stroke. Based on the Modified Rankin Scale classification (van Swieten, Koudstaal, Visser, Schouten, & van Gijn, 1988), 60 (29%) of the participants had mild strokes (scores of 1-2), 57 (27%) had moderate strokes (scores of 3), and 93 (44%) had severe strokes (scores of 4). Fifty-seven (27%) had a hemorrhagic stroke and 153 (73%) had an ischemic stroke. A total of 173 (84%) experienced paresis due to stroke, 81 (47%) on the right side, 83 (48%) on the left side, and 9 (5%) on both sides. Thirty-three (16%) participants did not experience weakness.

TBI severity was characterized by the Glasgow Coma Scale (GCS) obtained within 24 hours of injury. Sixty-seven (37%) had complicated-mild injury (GCS score of 13-15 and positive neuroimaging findings)(Williams, Levin, & Eisenberg, 1990), 16 (9%) had moderate injury (GCS score of 9-12), and 99 (54%) had severe injury (GCS score of 8 or lower) (Traumatic Brain Injury Model Systems National Data Center, 2006). Most injuries were caused by motor vehicle crashes ( $n = 97$ , 53%), falls ( $n = 50$ , 27%), gunshot wounds or other acts of violence ( $n = 21$ , 12%), sports injuries ( $n = 5$ , 3%), and other causes ( $n = 8$ , 4%). Cause of injury was unknown for one participant (1%).

Table 1 presents the demographic and clinical characteristics of the samples. Participants with stroke were older and closer to time-of-injury than participants with TBI. The TBI sample contained more men than women, whereas gender was equally distributed in the stroke sample. The samples were racially and ethnically diverse with more African-American participants in the stroke sample compared to the TBI sample. There were similar proportions of Hispanic participants in the two samples. Most participants reported having more than a high school level of education, and most were not working when assessed.

## Instruments

Participants completed the Cognition, Emotion, Sensory, and Motor batteries of the NIH Toolbox and additional neuropsychological tests. Study participation required 2 days for most participants, and participants received a stipend of US \$90 per day. Some participants required additional days of testing, and received a stipend of US \$20 per day. Participants provided informed consent in accordance with the local Institutional Review Board. We described the methods and procedures previously [MASKED].

### **NIH Toolbox of Neurological and Behavioral Functioning - Cognition Battery—**

The NIHTB-CB is a brief (<30 minute) set of cognitive tests for use with participants from 3-85 years (Gershon et al., 2010; Gershon, Wagster, et al., 2013). It assesses episodic memory (Bauer et al., 2013; Dikmen et al., 2014), reading (Gershon et al., 2014; Gershon, Slotkin, et al., 2013), vocabulary (Gershon et al., 2014; Gershon, Slotkin, et al., 2013), processing speed (Carlozzi et al., 2014), working memory (Tulsky et al., 2014; Tulsky et al., 2013), mental set shifting, and cognitive inhibition (Zelazo et al., 2013; Zelazo et al., 2014). Data collection was performed using the standardization versions of the NIHTB-CB that utilize computer administration with traditional computer hardware (e.g., two monitors, mouse input). Table 2 lists the domain-specific tests.

**Traditional neuropsychological tests**—We administered neuropsychological tests as criterion measures (Table 2), selecting the same tests that were used in the validation of the NIHTB-CB in order to evaluate model invariance (Mungas et al., 2014). These included tests of episodic memory (Brief Visuospatial Memory Test-Revised [BVM-T-R], and a three-trial immediate recall version of the Rey Auditory Verbal Learning Test [RAVLT]); tests of working memory (WAIS-IV Letter Number Sequencing [LNS]); tests of processing speed (WAIS-IV Digit Symbol Coding [DS] and WAIS-IV Symbol Search [SS]); tests of executive function (Wisconsin Card Sorting Test [WCST] and Delis-Kaplan Executive Functioning System [DKEFS]); and measures of crystallized cognition that assess language functioning (Peabody Picture Vocabulary Test-Revised [PPVT-R] and Wide-Range Achievement Test Reading [WRAT-IV]). Administration of the RAVLT and BVM-T-R included only the learning trials (i.e., not the delayed recall task) to replicate the administration procedures of the NIHTB-CB Picture Sequence Memory Test.

### Data Integrity

We took several steps to ensure that data collectors followed standardized test administration protocols [MASKED]. Following initial training on administration, each examiner practiced test administration for a minimum of 5 cases. Then, one author [MASKED] observed and certified them in a live testing session. We recertified examiners annually to ensure that examiners continued to administer tests in a standardized manner. Scoring of the NIHTB tests is done automatically by computer, and we monitored test scoring carefully for the additional neuropsychological tests. Every year one author [MASKED] reviewed 10 de-identified test protocols from each examiner, rescored the protocol, and provided feedback about deviations from standard procedures. In cases where scoring did not achieve 95% agreement, we retrained the examiner and reviewed scoring for 10 additional cases.

### Missing Data

Data were missing for 32.4% of the cases on at least one of the 17 cognitive measures; 3.8% of cases were missing 4 or more measures from the NIHTB-CB and 1.8% were missing 6 or more of the neuropsychological tests. We required data from at least 4 NIHTB-CB and 5 of the additional neuropsychological measures for inclusion in the CFA analysis. Using these criteria, 94.9% of cases were included in the analysis. Inclusion did not vary by the level of impairment within each injury type. We observed no differences regarding missing data related to age, years since injury, education, sex, race, or ethnicity. Individuals who were not employed or did not report employment status were excluded more often than those who were working,  $\chi^2(4) = 37.61, p < .001$ . Full information maximum likelihood (FIML) estimation was used to accommodate the remaining missing data. In contrast to imputation-based procedures, FIML uses all available data during model estimation without imputing values. FIML generally provides comparable results to multiple imputation; together, these missing-at-random procedures are state-of-the-art for missing data management (Enders, 2010).

## Confirmatory Factor Analysis

**Model development**—Based on general population studies of adults and adolescents described above, we expected a 5-factor model including: vocabulary, reading, episodic memory, working memory, and executive functioning/processing speed (Mungas et al., 2014). Mungas and colleagues rejected an alternative 6-factor model because the correlation between processing speed and executive function exceeded 1.0 and information criteria suggested slightly better fit with the 5-factor model (Mungas et al., 2014).

For this study, we tested the goodness-of-fit and parsimony of several alternative models from a 1-factor to a 6-factor model that Mungas et al. (2014) reported (Table 3). We tested a 6-factor model in which executive functioning and processing speed were two distinct factors, because processing speed has emerged as a distinct domain of functioning in factor analytic studies with other cognitive test batteries (Holdnack et al., 2011; Tulsky & Price, 2003). However, there is a strong relationship between the NIHTB-CB measures of executive functioning and processing speed measures because the NIHTB-CB executive function tests (Flanker and Dimensional Change Card Sorting test) are timed. The criterion executive functioning measures (WCST, DKEFS) are not timed. After identifying the best model(s) in the combined brain injury sample, we repeated CFAs to determine if the models fit the TBI and stroke samples separately, as well as a sample of individuals with severe injury.

**Goodness of fit**—To determine the best measurement model, we compared the multiple models outlined in Table 3. We compared the fit statistics for each model with the previous, less complex model in the sequence (e.g., a two-factor solution vs. the one-factor model; a three-factor solution vs. the two-factor solution). Modification indices were reviewed for possible cross-loadings of NIHTB-CB tests that may improve goodness-of-fit. As in Mungas et al. (2014), we allowed the error terms for tests with similar methods to correlate – specifically, the error terms of (a) the criterion WAIS-IV Coding and WAIS-IV Symbol Search tests, and (b) the NIHTB-CB DCCS and Flanker tests.

We evaluated these models using a variety of goodness-of-fit indices. Criteria based on the chi-square statistic (i.e., likelihood ratio chi-square statistic) have been used to evaluate model fit. However, the chi-square statistic is sensitive to large samples. Hence, we used alternative fit indices because they are less sensitive to large samples (Bollen, 1989; Bollen & Long, 1993; Byrne, 2001; Hu & Bentler, 1999; Marsh, 1988; Schumacker & Lomax, 2004; Tanaka, 1993; Thompson, 2000). These alternatives include the comparative fit index (CFI)(Bentler, 1990), an index that compares the proposed model to a baseline model; the Tucker–Lewis fit index (TLI)(Hu & Bentler, 1999; Tucker, 1973), a comparative fit index that adjusts the degrees of freedom in the model; the root mean square error of approximation (RMSEA)(Browne & Cudeck, 1993; Hu & Bentler, 1999; J. H. Steiger, 1990), which compensates for the effect of model complexity by dividing the *F* statistic by the degrees of freedom; and the standardized root mean square residual (SRMR) index, a standardized measure of the degree of reproduction of the covariance matrix from the model estimates. Finally, the Bayesian information criterion (BIC)(Schwartz, 1978) helps evaluate the evidence from several models and favors parsimonious solutions (Raftery, 1993).



Excellent model fit is defined by CFI values of .95 or higher (Bentler, 1990; Hu & Bentler, 1999), TLI values of .95 or higher (Hu & Bentler, 1999), and SRMR and RMSEA values of .05 or lower (Hu & Bentler, 1999). Good to adequate fit is defined by CFI and TLI values of .90 and RMSEA and SRMR values of .08 or lower (Browne & Cudeck, 1993; MacCallum, Browne, & Sugawara, 1996; J. Steiger, 2007). Larger values indicate better fit for the CFI and TLI and smaller values for SRMR and RMSEA. Raftery suggested that a BIC difference of 10 or greater provides very strong evidence for model preference, while a BIC difference of 5-9 offers strong evidence (Raftery, 1993). All models were estimated in the R package lavaan 0.5-22 (Rosseel, 2012).

## Results

### Joint Factor Analysis

Table 4 presents the results of the confirmatory factor analyses. The best fitting model in the general population sample (Mungas et al., 2014) (model 5b: vocabulary, reading, episodic memory, working memory, and a combined executive function/processing speed factor) had good fit in this sample of individuals with TBI and Stroke,  $\chi^2(107) = 288.03$ ,  $\chi^2/df = 2.69$ , CFI = .96, TLI = .95, RMSEA = .07, SRMR = .04. According to the BIC, this model resulted in the second-lowest BIC value (660.92); the 6-factor model – in which processing speed and executive function are split into two distinct factors – had the lowest BIC (660.86). Although technically superior in fit, the 6-factor model resulted in a correlation between the processing speed factor and the executive factor of .98; thus, we regard Model 5b as the most reasonable representation of the data in this sample, which replicates the structure reported by Mungas et al. (2014).

Figure 1 displays the factor correlations and standardized factor loadings for model 5b. The factor correlations were high to very high with the exception of the correlations between reading and processing speed/executive function (.44), reading and episodic memory (.45), vocabulary and processing speed/executive function (.46), and vocabulary and episodic memory (.44). The highest correlations were between the episodic and working memory factors (.83), the episodic memory and processing speed/executive function factors (.85), and the working memory and processing speed/executive function factors (.87). The NIHTB measures had the strongest loading for reading (Oral Reading = .96) and vocabulary (Picture Vocabulary = .96) and the lowest for processing speed/executive function (DCCS = .75). The error covariance between the two NIHTB-CB executive function measures (DCCS and Flanker) was positive and statistically significant (raw estimate = .14,  $SE = .06$ ,  $p = .02$ ). Likewise, the error covariance between the two processing speed criterion measures (WAIS-IV Coding and WAIS-IV Symbol Search) was positive and significantly different from zero (raw estimate = 20.96,  $SE = 3.77$ ,  $p < .01$ ). Evidently, shared method variance emerged between these two pairs of tests as in Mungas et al. (2014). Modification indices revealed several cross-loadings and error covariances that would significantly improve model fit if estimated. However, the two highest values were associated with pathways that did not correspond to those identified by Mungas and colleagues (2014); additionally, none of the modification indices were clear outlying values. Therefore, we did not estimate additional paths to improve the fit of Model 5b.

## CFA Results by Injury Type and Severity

We conducted additional analyses to learn if the best fitting model in the total sample also fit the TBI and stroke subsamples. Severe injuries are more likely to be associated with global cognitive impairment, which may interfere with the ability to differentiate specific cognitive deficits. We therefore repeated the same procedures described for the total sample with these three subsamples.

**TBI sample**—The confirmatory factor analysis that included both the NIHTB-CB subtests and traditional tests showed acceptable fit in the TBI sample. Fit indices for Model 5b were:  $\chi^2(107) = 224.72$ ,  $\chi^2/df = 2.10$ , CFI = .94, TLI = .93, RMSEA = .08, SRMR = .05. This model fit the data better than all other models, including the 6-factor model, based on the BIC (549.75 vs. 568.83). The second-lowest BIC value was found in Model 4c (552.32). Model 4c mirrors Model 5b with the exception of the episodic and working memory factors, which we reduced to a single factor in Model 4c. The small difference between these models on all fit indices and the strong correlation between the episodic memory and working memory factors in Model 5b (.89) makes it difficult to conclude which structure provides a better fit.

**Stroke sample**—Similar to the overall sample, the 6-factor model fit slightly better (BIC = 535.99) compared to the 5-factor (Model 5b) model (BIC = 536.34) in the Stroke subsample. The correlation between the processing speed and executive function factors was close to 1 (.98), suggesting that Model 5b is a more parsimonious and plausible solution despite having the second-lowest BIC. Other fit indices for Model 5b were in acceptable ranges:  $\chi^2(107) = 203.18$ ,  $\chi^2/df = 1.90$ , CFI = .96, TLI = .95, RMSEA = .07, SRMR = .05. Modification indices did not identify model improvements that would result from the estimation of additional cross-loadings or error covariances.

**Injury severity**—Model 5b also fit the data from the severe injury subsample adequately,  $\chi^2(107) = 200.12$ ,  $\chi^2/df = 1.87$ , CFI = .96, TLI = .95, RMSEA = .07, SRMR = .04. Model 5b also had the lowest BIC value (526.92), followed by Model 6 (533.76) and Model 4b (542.54). Parameter estimates from Model 5b were comparable to those observed in the overall sample, and no modification indices suggested salient forms of model misspecification. Therefore, despite having greater levels of cognitive deficits following TBI or Stroke, individuals with more severe injuries appear to adhere to the same 5-factor structure underlying the combined set of NIHTB-CB and cognitive criterion measures.

## Discussion

We evaluated the extent to which Mungas et al.'s model of cognition (2014) in the general adult population fits individuals with brain injury due to stroke or TBI. We observed strong support for Mungas' general population model in which executive functioning and processing speed tests load on a single factor. Since the DCCS and Flanker tests are scored based upon a combination of speed and accuracy, much of the variance of these executive functioning tests reflects performance speed. The fit to model 5b (which suggests combining



processing speed and executive functioning on a single factor) is supported by the data in this brain injury sample as well.

We also observed support for model 6f (i.e., vocabulary, reading, episodic memory, working memory, executive functioning, and processing speed), which suggests that the executive functioning and processing speed tests split onto two distinct factors. Although most of the fit statistics (e.g., CFI, TLI, RMSEA, and BIC) improve with a 6-factor solution, there is a high correlation between the executive functioning and processing speed factors; the magnitude of this correlation coefficient is too high to justify retaining executive functioning and processing speed as distinct factors. Because the NIHTB-CB executive functioning tests are scored based on accuracy *and* speed, the high correlation between these factors is expected. In fact, the emphasis of timed performance in the scores on Flanker Inhibitory Control and DCCS creates *de facto* processing speed tests. For this reason, we do not think the results support a 6-factor solution where executive functioning and processing speed are distinct factors.

Picture Vocabulary and Oral Reading tests, measures of crystallized cognition, load on separate factors in Model 5B. A four-factor model with reading and vocabulary tests loading on a single factor of crystallized language demonstrated acceptable model fit; however, we obtained better model fit with these tests loading on separate but correlated factors of reading and vocabulary (models 5b and 6f). The reading and vocabulary tests tap crystallized cognition that is less affected by brain injury than is fluid cognition (Akshoomoff et al., 2013; Heaton et al., 2014; Larrabee, Lergen, & Levin, 1985); thus, it is not surprising that these tests correlate highly. The correlation of reading with vocabulary is a fairly modest 0.71, supporting our decision to retain distinct factors.

The NIHTB-CB does not include measures of visual-perceptual or visual-spatial processing. Thus, we cannot determine the degree to which spatial functioning may affect performance on tests such as PSM, which requires processing of visual-perceptual information and spatial sequencing. Motor tests were not included in this CFA either; the influence of motor skills on reaction time tasks cannot be evaluated. Further research is required to elucidate a fuller spectrum of cognitive and other skills that influence performance on the NIHTB-CB.

In summary, the factor structure of the NIHTB-CB in a sample of individuals with acquired brain injury is similar to the structure reported by Mungas and colleagues in a general population sample. The results provide strong support for the construct validity of the NIHTB-CB for adults with acquired brain injury. The large and diverse clinical populations with brain injury are strengths of this study. Future studies of utilizing the NIHTB-CB in adults with brain injury will help marshal further evidence of construct validity and help advance use of these tests in research and practice.

## References

- Akshoomoff N, Beaumont JL, Bauer PJ, Dikmen SS, Gershon RC, Mungas D, Heaton RK. VIII. NIH Toolbox Cognition Battery (CB): composite scores of crystallized, fluid, and overall cognition. *Monographs of the Society for Research in Child Development*. 2013; 78(4):119–132. DOI: 10.1111/mono.12038 [PubMed: 23952206]

- Bauer PJ, Dikmen SS, Heaton RK, Mungas D, Slotkin J, Beaumont JL. III. NIH Toolbox Cognition Battery (CB): measuring episodic memory. *Monographs of the Society for Research in Child Development*. 2013; 78(4):34–48. DOI: 10.1111/mono.12033 [PubMed: 23952201]
- Benson N, Hulac DM, Kranzler JH. Independent examination of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV): what does the WAIS-IV measure? *Psychological Assessment*. 2010; 22(1):121–130. DOI: 10.1037/a0017767 [PubMed: 20230158]
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107(2):238–246. [PubMed: 2320703]
- Bodin D, Pardini DA, Burns TG, Stevens AB. Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology*. 2009; 15(5):417–424. DOI: 10.1080/09297040802603661 [PubMed: 19132580]
- Bollen, KA. *Structural equations with latent variables*. New York: Wiley; 1989.
- Bollen, KA., Long, JS. *Testing Structural Equation Models*. Newbury Park, CA: Sage; 1993.
- Bowde SC, Ritter AJ, Carstairs JR, Shores EA, Pead J, Greeley JD, Clifford CC. Factorial invariance for combined Wechsler Adult Intelligence Scale-Revised and Wechsler Memory Scale-Revised scores in a sample of clients with alcohol dependency. *The Clinical Neuropsychologist*. 2001; 15(1): 69–80. DOI: 10.1076/clin.15.1.69.1910 [PubMed: 11778580]
- Bowden SC, Carstairs JR, Shores EA. Confirmatory Factor Analysis of Combined Wechsler Adult Intelligence Scale-Revised and Wechsler Memory Scale-Revised Scores in a Healthy Community Sample. *Psychological Assessment*. 1999; 11(3):339–344.
- Bowden SC, Lissner D, McCarthy KA, Weiss LG, Holdnack JA. Metric and structural equivalence of core cognitive abilities measured with the Wechsler Adult Intelligence Scale-III in the United States and Australia. *Journal of Clinical and Experimental Neuropsychology*. 2007; 29(7):768–780. DOI: 10.1080/13803390601028027 [PubMed: 17896201]
- Bowden SC, Saklofske DH, Weiss LG. Augmenting the core battery with supplementary subtests: Wechsler adult intelligence scale-IV measurement invariance across the United States and Canada. *Assessment*. 2011; 18(2):133–140. DOI: 10.1177/1073191110381717 [PubMed: 20826685]
- Bowden SC, Weiss LG, Holdnack JA, Bardenhagen FJ, Cook MJ. Equivalence of a measurement model of cognitive abilities in U.S. standardization and Australian neuroscience samples. *Assessment*. 2008; 15(2):132–144. DOI: 10.1177/1073191107309345 [PubMed: 18463404]
- Browne, MW., Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA., Long, JA., editors. *Testing Structural Equation Models*. Sage Publications; Thousand Oaks: 1993. p. 136-162.
- Byrne BM. *Structural Equation Modeling: Perspectives on the Present and the Future*. *International Journal of Testing*. 2001; 1:327–334.
- Carlozzi NE, Tulsky DS, Chiaravalloti ND, Beaumont JL, Weintraub S, Conway K, Gershon RC. NIH Toolbox Cognitive Battery (NIHTB-CB): The NIHTB Pattern Comparison Processing Speed Test. *Journal of the International Neuropsychological Society*. 2014; 20:630–641. [PubMed: 24960594]
- Carlozzi NE, Tulsky DS, Wolf T, Goodnight S, Heaton R, Casaletto K, Heinemann AW. Construct validity of the NIH Toolbox Cognition Battery in Individuals with Stroke. (Under Review).
- Casaletto KB, Umlauf A, Beaumont J, Gershon R, Slotkin J, Akshoomoff N, Heaton RK. Demographically Corrected Normative Standards for the English Version of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*. 2015; 21(5):378–391. DOI: 10.1017/S1355617715000351 [PubMed: 26030001]
- Cohen ML, Tulsky DS, Holdnack JA, Carlozzi NE, Wong A, Magasi S, Heinemann AW. Cognition among Community-Dwelling Individuals with Spinal Cord Injury. (Under Review).
- Dickinson D, Iannone VN, Gold JM. Factor structure of the Wechsler Adult Intelligence Scale-III in schizophrenia. *Assessment*. 2002; 9(2):171–180. [PubMed: 12066832]
- Dikmen SS, Bauer PJ, Weintraub S, Mungas D, Slotkin J, Beaumont JL, Heaton RK. Measuring episodic memory across the lifespan: NIH Toolbox Picture Sequence Memory Test. *Journal of the International Neuropsychological Society*. 2014; 20(6):611–619. DOI: 10.1017/S1355617714000460 [PubMed: 24960230]
- Enders, CK. *Applied Missing Data Analysis*. New York: Guilford Press; 2010.

- Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, Wagster MV. Assessment of neurological and behavioural function: the NIH Toolbox. *Lancet Neurology*. 2010; 9(2):138–139. DOI: 10.1016/S1474-4422(09)70335-7 [PubMed: 20129161]
- Gershon RC, Cook KF, Mungas D, Manly JJ, Slotkin J, Beaumont JL, Weintraub S. Language measures of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*. 2014; 20(6):642–651. DOI: 10.1017/S1355617714000411 [PubMed: 24960128]
- Gershon RC, Slotkin J, Manly JJ, Blitz DL, Beaumont JL, Schnipke D, Weintraub S. IV. NIH Toolbox Cognition Battery (CB): measuring language (vocabulary comprehension and reading decoding). *Monographs of the Society for Research in Child Development*. 2013; 78(4):49–69. DOI: 10.1111/mono.12034 [PubMed: 23952202]
- Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH Toolbox for Assessment of Neurological and Behavioral Function. *Neurology*. 2013; 80(Suppl 3):S2–S6. [PubMed: 23479538]
- Heaton RK, Akshoomoff N, Tulsky D, Mungas D, Weintraub S, Dikmen S, Gershon R. Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *Journal of the International Neuropsychological Society*. 2014; 20(6):588–598. DOI: 10.1017/S1355617714000241 [PubMed: 24960398]
- Holdnack JA, Xiaobin Z, Larrabee GJ, Millis SR, Salthouse TA. Confirmatory factor analysis of the WAIS-IV/WMS-IV. *Assessment*. 2011; 18(2):178–191. DOI: 10.1177/1073191110393106 [PubMed: 21208975]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6(1):1–55. DOI: 10.1080/10705519909540118
- Larrabee GJ, Lergen JW, Levin HS. Sensitivity of age-decline resistant (“hold”) WAIS subtests to Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*. 1985; 7(5):497–504. DOI: 10.1080/01688638508401281 [PubMed: 4066906]
- MacCallum R, Browne M, Sugawara H. Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods*. 1996; 1(2):130–149.
- Marsh H, Balla JR, McDonald RP. Goodness-of-fit Indices in Confirmatory Factor Analysis: Effects of Sample Size. *Psychological Bulletin*. 1988; 103:391–411.
- Mungas D, Heaton R, Tulsky D, Zelazo PD, Slotkin J, Blitz D, Gershon R. Factor structure, convergent validity, and discriminant validity of the NIH Toolbox Cognitive Health Battery (NIHTB-CHB) in adults. *Journal of the International Neuropsychological Society*. 2014; 20(6):579–587. DOI: 10.1017/S1355617714000307 [PubMed: 24960474]
- Mungas D, Widaman K, Zelazo PD, Tulsky D, Heaton RK, Slotkin J, Gershon RC. VII. NIH Toolbox Cognition Battery (CB): factor structure for 3 to 15 year olds. *Monographs of the Society for Research in Child Development*. 2013; 78(4):103–118. DOI: 10.1111/mono.12037 [PubMed: 23952205]
- Niileksela CR, Reynolds MR, Kaufman AS. An alternative Cattell-Horn-Carroll (CHC) factor structure of the WAIS-IV: age invariance of an alternative model for ages 70-90. *Psychological Assessment*. 2013; 25(2):391–404. DOI: 10.1037/a0031175 [PubMed: 23244639]
- Raftery, A. Bayesian Model Selection in Structural Equation Models. In: K, A.B, J., S, Long, editors. *Testing Structural Equation Models*. Newbury Park, CA: Sage; 1993. p. 163-180.
- Rosseel Y. lavaan: An R package for structural equation modeling. *Journal of Statistical Software*. 2012; 48:1–36.
- Schumacker, R., Lomax, R. *A Beginner's Guide to Structural Equation Modeling*. 2nd. Mahwah, NJ: Erlbaum; 2004.
- Schwartz G. Estimating the Dimension of a Model. *Annals of Statistics*. 1978; 6:461–464.
- Steiger J. Understanding the Limitations of Global Fit Assessment in Structural Equation Modeling. *Personality and Individual Differences*. 2007; 42(5):893–898.
- Steiger JH. Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*. 1990; 25(2):173–180. DOI: 10.1207/s15327906mbr2502\_4 [PubMed: 26794479]
- Stevens, J. *Applied Multivariate Statistics for the Social Sciences*. Mahwah, NJ: Erlbaum; 1996.

- Tanaka, JS. Multifaceted Conceptions of Fit in Structural Equation Models. In: K, A.B, J., S, Long, editors. *Testing Structural Equation Models*. Newbury Park, CA: Sage; 1993. p. 10-12.
- Taub GE, McGrew KS, Witta EL. A confirmatory analysis of the factor structure and cross-age invariance of the Wechsler Adult Intelligence Scale-Third Edition. *Psychological Assessment*. 2004; 16(1):85–89. DOI: 10.1037/1040-3590.16.1.85 [PubMed: 15023096]
- Thompson, B. Ten Commandments of Structural Equation Modeling. In: G, LG., Yarnold, PR., editors. *Reading and Understanding MORE Multivariate Statistics*. Washington, DC: American Psychological Association; 2000. p. 261-283.
- Traumatic Brain Injury Model Systems National Data Center. Traumatic Brain Injury Model Systems National Data Base Inclusion Criteria. 2006. Retrieved from <http://www.tbindsc.org/Documents/2010%20TBIMS%20Slide%20Presentation.pdf>
- Tucker LR, L C. A Reliability Coefficient for Maximum Likelihood Factor Analysis. *Psychometrika*. 1973; 38:1–10.
- Tulsky DS, Carlozzi N, Chiaravalloti ND, Beaumont JL, Kisala PA, Mungas D, Gershon R. NIH Toolbox Cognition Battery (NIHTB-CB): list sorting test to measure working memory. *Journal of the International Neuropsychological Society*. 2014; 20(6):599–610. DOI: 10.1017/S135561771400040X [PubMed: 24959983]
- Tulsky DS, Carlozzi NE, Chevalier N, Espy KA, Beaumont JL, Mungas D. V. NIH Toolbox Cognition Battery (CB): measuring working memory. *Monographs of the Society for Research in Child Development*. 2013; 78(4):70–87. DOI: 10.1111/mono.12035 [PubMed: 23952203]
- Tulsky DS, Carlozzi NE, Holdnack JA, Heaton RK, Wong A, Goldsmith A, Heinemann AW. Using the NIH Toolbox Cognition Battery (NIHTB-CB) in Individuals with a Traumatic Brain Injury. (Under Review).
- Tulsky DS, Heinemann AW. Establishing the clinical utility and construct validity of the NIH Toolbox Cognitive Battery in Individuals with Disabilities. (Under Review).
- Tulsky DS, Price LR. The joint WAIS-III and WMS-III factor structure: development and cross-validation of a six-factor model of cognitive functioning. *Psychological Assessment*. 2003; 15(2): 149–162. [PubMed: 12847775]
- van der Heijden P, Donders J. A confirmatory factor analysis of the WAIS-III in patients with traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*. 2003; 25(1):59–65. DOI: 10.1076/jcen.25.1.59.13627 [PubMed: 12607172]
- van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke*. 1988; 19(5):604–607. [PubMed: 3363593]
- Ward LC, Bergman MA, Hebert KR. WAIS-IV subtest covariance structure: conceptual and statistical considerations. *Psychological Assessment*. 2012; 24(2):328–340. DOI: 10.1037/a0025614 [PubMed: 21928907]
- Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, Gershon RC. Cognition assessment using the NIH Toolbox. *Neurology*. 2013; 80(11 Suppl 3):S54–64. DOI: 10.1212/WNL.0b013e3182872ded [PubMed: 23479546]
- Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Slotkin J, Gershon R. The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: validation in an adult sample. *Journal of the International Neuropsychological Society*. 2014; 20(6):567–578. DOI: 10.1017/S1355617714000320 [PubMed: 24959840]
- Weiss LG, Keith TZ, Zhu J, Chen H. WAIS-IV Clinical Validation of the Four- and Five-Factor Interpretive Approaches. *Journal of Psychoeducational Assessment*. 2013a; 31:94–113.
- Weiss LG, Keith TZ, Zhu J, Chen H. WISC-IV Clinical Validation of the Four- and Five-factor Interpretive Approaches. *Journal of Psychoeducational Assessment*. 2013b; 31:114–131.
- Williams DH, Levin HS, Eisenberg HM. Mild Head-Injury Classification. *Neurosurgery*. 1990; 27(3): 422–428. [PubMed: 2234336]
- Zelazo PD, Andersen J, Richler J, Wallner-Allen K, Beaumont J, Weintraub S. NIH Toolbox Cognitive Function Battery (CFB): Measuring Executive Function and Attention. *Monographs of the Society for Research in Child Development*. 2013; 78(4):16–33. [PubMed: 23952200]
- Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont JL, Conway KP, Weintraub S. NIH Toolbox Cognition Battery (CB): validation of executive function measures in adults. *Journal of*

the International Neuropsychological Society. 2014; 20(6):620–629. DOI: 10.1017/S1355617714000472 [PubMed: 24960301]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Impact**

- The NIH Toolbox Cognition Battery was developed for and validated in the general population. It has great potential for clinical and research applications with individuals with acquired brain injury, but requires validation in these groups to establish its clinical utility.
- This study provides evidence of construct validity of the NIHTB-CB by replicating the factor structure in a sample of adults with acquired brain injury.
- The American Psychological Association guidelines state that new tests must demonstrate construct validity prior to use in practice or research. The results of this study support the use of the NIHTB-CB for clinical research and practice with adults with acquired brain injury.



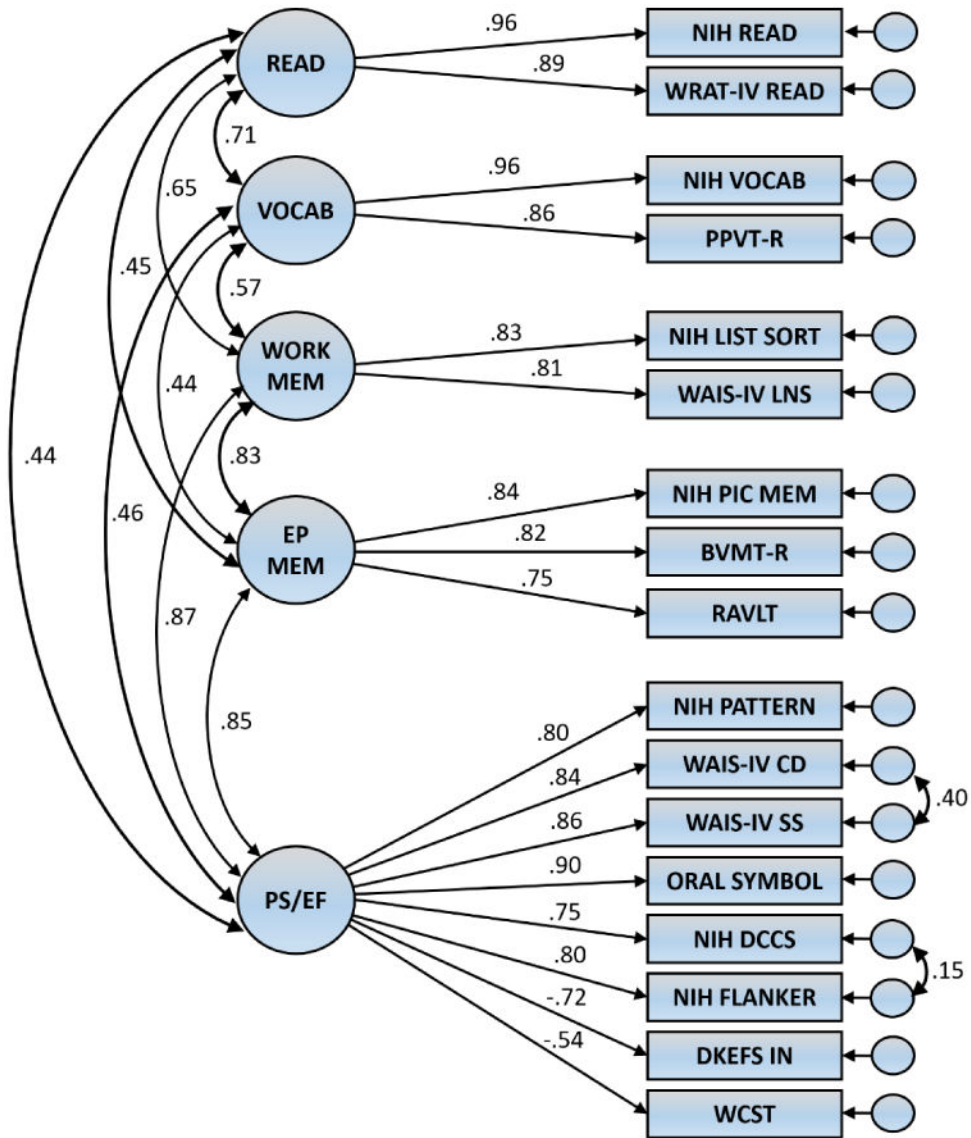


Figure 1. Standardized Factor Loadings, Factor Correlations, and Error Covariances for Joint Factor Analysis Model 5fb: Language, Episodic Memory, Working Memory, Executive Function, and Processing Speed

**Table 1**  
**TBI and Stroke Group Demographic and Injury Characteristics**

Variable	TBI (N=182)	Stroke (N=210)
Age(Years)		
M (SD)	39.1 (17.0)	56.2 (12.9)
Time Since Injury (years)		
M(SD)	6.0 (5.5)	2.8 (2.5)
Sex (%)		
Male	63.7	50.0
Female	36.3	50.0
Race (%)		
Caucasian	73.6	42.9
African American	15.9	49.5
Other	10.5	7.6
Ethnicity (%)		
Not Hispanic or Latino	92.3	94.2
Hispanic or Latino	7.1	4.8
Not Provided	0.6	1.0
Education (%)		
Less than 12 years	13.8	13.3
12 years	20.4	20.0
13-15 years	34.8	38.1
16 or more years	30.9	28.6
Education (Years)		
M (SD)	13.7 (2.4)	13.7 (2.6)
Injury Severity (%)		
Complicated-Mild <sup>a</sup>	36.8	28.6
Moderate	8.8	27.1
Severe	54.4	44.3
Work Status (%)		
Full-Time	19.8	17.6
Part-Time	23.1	13.8
Volunteer	0.6	0.5
Not Employed	51.6	64.8
Unknown	4.9	3.3

Note:

<sup>a</sup> as defined by Williams et al. (1990)

**Table 2**  
**Measures and Associated Domains**

Measure	Associated Domain
Picture Vocabulary *	Vocabulary, Language, Crystallized/Global
Oral Reading *	Reading, Language, Crystallized, Global
Picture Sequence Memory *	Episodic Memory, Fluid, Global
List Sorting *	Working Memory, Executive, Fluid, Global
Flanker *	Executive, Fluid, Global
DCCS *	Executive, Fluid, Global
Pattern Comparison *	Processing Speed, Executive, Fluid, Global
Oral Symbol Digit **	Processing Speed, Executive, Fluid, Global
RAVLT (NIHTB-CB) **	Episodic Memory, Fluid, Global
PPVT-R	Vocabulary, Language, Crystallized/Global
WRAT-IV Reading	Reading, Language, Crystallized, Global
BVMT-R	Episodic Memory, Fluid, Global
WAIS-IV LNS	Working Memory, Executive, Fluid, Global
WAIS-IV DS	Processing Speed, Fluid, Global
WAIS-IV SS	Processing Speed, Fluid, Global
WCST Total Errors	Executive, Fluid, Global
DKEFS CWIT IN	Executive, Fluid, Global

\* Test is part of the NIHTB-CB.

\*\* Test is optional NIHTB-CB Toolbox test.

Note: "Associated domains" indicates primary cognitive ability measured by test and secondary higher order domains. DCCS = Dimensional Change Card Sort, PPVT-R = Peabody Picture Vocabulary Test-Revised Edition, WRAT-IV Reading = Wide Range Achievement Test-4th Edition Reading, RAVLT = Rey Auditory Verbal Learning Test, BVMT-R = Brief Visual Memory Test-Revised Edition, WAIS-IV LNS = Wechsler Adult Intelligence Scale-3rd edition Letter-Number Sequencing, WAIS-IV CD = Wechsler Adult Intelligence Scale-3rd edition Coding, WAIS-IV SS = Wechsler Adult Intelligence Scale-3rd edition Symbol Search, WCST Total Errors = Wisconsin Card Sorting Test Total Errors, DKEFS CWIT IN = Delis-Kaplan Executive Functioning System Color Word Interference Test Inhibition Condition

**Table 3**  
**Cognitive Models Evaluated with Joint NIH Toolbox and Established Neuropsychological Tests CFA**

---

1f– Global Cognition
2fa– Crystallized, Fluid
2fb– Episodic Memory/Working Memory, Non-Memory
3fa– Language, Episodic Memory/Working Memory, Executive/Speed
3fb – Language, Episodic Memory, Working Memory/Executive/Speed
4fa – Language, Episodic Memory, Working Memory, Executive/Speed
4fb– Vocabulary, Reading, Episodic Memory, Working Memory/Executive/Speed
4fc– Vocabulary, Reading, Episodic Memory/Working Memory, Executive/Speed
5fa– Language, Episodic Memory, Working Memory, Executive, Speed
5fb– Vocabulary, Reading, Episodic Memory, Working Memory, Executive/Speed
6f – Vocabulary, Reading, Episodic Memory, Working Memory, Executive, Speed

---

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4**  
**CFA Model Fit Statistics for Joint NIH Toolbox and Established Neuropsychological Tests**

Model	$\chi^2$	df	$\chi^2/df$	CFI	TLI	RMSEA	SRMR	BIC
1f	1289.13	117	11.02	0.74	0.70	0.16	0.10	1602.83
2a	652.06	116	5.62	0.88	0.86	0.11	0.07	971.69
2b	1238.22	116	10.67	0.76	0.71	0.16	0.10	1557.85
3a	553.25	114	4.85	0.90	0.89	0.10	0.06	884.71
3b	573.06	114	5.03	0.90	0.88	0.10	0.06	904.52
4a	478.39	111	4.31	0.92	0.90	0.09	0.06	827.60
4b	381.49	111	3.44	0.94	0.93	0.08	0.05	730.71
4c	361.40	111	3.26	0.94	0.93	0.08	0.05	710.61
5a	447.71	107	4.18	0.93	0.91	0.09	0.05	820.60
5b	288.03	107	2.69	0.96	0.95	0.07	0.04	660.92
6f	258.37	102	2.53	0.97	0.96	0.06	0.04	660.86