

# UCSF

## UC San Francisco Previously Published Works

### Title

Validity of the Brief Test of Adult Cognition by Telephone in Level 1 Trauma Center Patients Six Months Post-Traumatic Brain Injury: A TRACK-TBI Study

### Permalink

<https://escholarship.org/uc/item/0k71k81r>

### Journal

Journal of Neurotrauma, 38(8)

### ISSN

0897-7151

### Authors

Nelson, Lindsay D  
Barber, Jason K  
Temkin, Nancy R  
[et al.](#)

### Publication Date

2021-04-15

### DOI

10.1089/neu.2020.7295

Peer reviewed

# Validity of the Brief Test of Adult Cognition by Telephone in Level 1 Trauma Center Patients Six Months Post-Traumatic Brain Injury: A TRACK-TBI Study

Lindsay D. Nelson,<sup>1</sup> Jason K. Barber,<sup>2</sup> Nancy R. Temkin,<sup>3</sup> Kristen Dams-O'Connor,<sup>4</sup> Sureyya Dikmen,<sup>5</sup> Joseph T. Giacino,<sup>6</sup> Mark D. Kramer,<sup>7</sup> Harvey S. Levin,<sup>8</sup> Michael A. McCrea,<sup>1</sup> John Whyte,<sup>9</sup> Yelena G. Bodien,<sup>10</sup> John K. Yue,<sup>11</sup> Geoffrey T. Manley<sup>11</sup>; and the TRACK-TBI Investigators\*

## Abstract

Our objective was to examine the construct validity of the Brief Test of Adult Cognition by Telephone (BTACT) and its relationship to traumatic brain injury (TBI) of differing severities. Data were analyzed on 1422 patients with TBI and 170 orthopedic trauma controls (OTC) from the multi-center Transforming Research and Clinical Knowledge in TBI (TRACK-TBI) study. Participants were assessed at 6 months post-injury with the BTACT and an in-person neuropsychological battery. We examined the BTACT's factor structure, factorial group invariance, convergent and discriminant validity, and relationship to TBI and TBI severity. Confirmatory factor analysis supported both a 1-factor model and a 2-factor model comprising correlated Episodic Memory and Executive Function (EF) factors. Both models demonstrated strict invariance across TBI severity and OTC groups. Correlations between BTACT and criterion measures suggested that the BTACT memory indices predominantly reflect verbal episodic memory, whereas the BTACT EF factor correlated with a diverse range of cognitive tests. Although the EF factor and other BTACT indices showed significant relationships with TBI and TBI severity, some group effect sizes were larger for more comprehensive in-person cognitive tests than the BTACT. The BTACT is a promising, brief, phone-based cognitive screening tool for patients with TBI. Although the BTACT's memory items appear to index verbal Episodic Memory, items that purport to assess EFs may reflect a broader array of cognitive domains. The sensitivity of the BTACT to TBI severity is lower than domain-specific neuropsychological measures, suggesting it should not be used as a substitute for comprehensive, in-person cognitive testing at 6 months post-TBI.

**Keywords:** Brief Test of Adult Cognition by Telephone; BTACT; phone-based cognitive assessment; telemedicine; traumatic brain injury

## Introduction

ASSESSING COGNITIVE FUNCTIONING is important both for research on traumatic brain injury (TBI) and clinical care of patients with TBI.<sup>1</sup> Traditional neuropsychological testing requires a significant time commitment and examiner expertise, typically at an in-person appointment, factors that hinder the feasibility of

performing cognitive assessments in large-scale epidemiological studies and with patients who have mobility or transportation issues. Unsurprisingly, studies tend to report more missing data on cognitive performance measures as compared with questionnaire or interview-based assessments that are more adaptable to other modes of assessment.<sup>2,3</sup> Likewise, in clinical settings there is increasing interest in offering remote (telemedicine) service delivery

<sup>1</sup>Departments of Neurosurgery and Neurology, Medical College of Wisconsin, Milwaukee, Wisconsin, USA.

<sup>2</sup>Department of Neurological Surgery, <sup>3</sup>Departments of Neurological Surgery and Biostatistics, <sup>5</sup>Department of Rehabilitation Medicine, University of Washington, Seattle, Washington, USA.

<sup>4</sup>Departments of Rehabilitation and Human Performance, Neurology, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

<sup>6</sup>Department of Physical Medicine and Rehabilitation, Harvard Medical School; Spaulding Rehabilitation Hospital, Charlestown, Massachusetts, USA.

<sup>7</sup>Independent Consultant, Minneapolis, Minnesota, USA.

<sup>8</sup>Department of Physical Medicine and Rehabilitation, Baylor College of Medicine, Houston, Texas, USA.

<sup>9</sup>Moss Rehabilitation Research Institute, Elkins Park, Pennsylvania, USA.

<sup>10</sup>Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston Massachusetts, USA; Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Boston Massachusetts, USA.

<sup>11</sup>Department of Neurosurgery, University of California, San Francisco, San Francisco, California, USA.

\*The TRACK-TBI Investigators may be found at the end of this article.

for patients in rural areas or who otherwise cannot readily attend in-person appointments.<sup>4-6</sup> Phone-based cognitive assessment would increase access to neuropsychological services but has not been validated for the TBI population. This study leveraged an ongoing multi-center prospective study of civilian TBI to investigate the validity of a phone-based cognitive screening tool—the Brief Test of Adult Cognition by Telephone (BTACT)—to assess cognitive function 6 months post-TBI.

The BTACT is a cognitive screening measure developed for the epidemiological Midlife in the United States (MIDUS) study.<sup>7</sup> Unlike other phone-based cognitive tests that emphasize screening for dementia, the BTACT was intended to be sensitive to individual differences in normal aging. The instrument has a number of strengths that make it promising for use with the diverse TBI patient population, including brevity (about a 20-min administration time), assessment of cognitive domains commonly affected by TBI (i.e., learning and memory, executive functioning), and its adaptation of several widely used and well-validated traditional neuropsychological tests to the phone mode of assessment.<sup>8</sup> Normative data are available from the MIDUS study samples for a wide age range (32–84 years), and methodologically strong work has been performed to support the instrument's factor structure, test-retest reliability, alternate forms reliability, and concordance between in-person and phone administration.<sup>8,9</sup>

Limited data are available, however, verifying the extent to which the BTACT measures the cognitive domains it is purported to measure (i.e., concept of interest [CoI])<sup>10</sup> or demonstrating its ability to detect the cognitive sequelae of TBI (i.e., purpose of use [PoU]).<sup>11</sup> Prior studies in the general population and TBI samples have relied on the BTACT as the sole neuropsychological assessment, rendering its concurrent validity unknown. One study reported quite modest associations between some BTACT subtests and traditional neuropsychological tests performed 1–2 years apart (e.g.,  $r=0.23$  to  $0.33$  between BTACT memory indices and those of a story memory test), yet it is unclear the degree to which the weak associations reported were due to limitations in the instrument's construct validity versus expected instability in cognitive performance over a long time interval.<sup>8</sup> One large-scale study of patients with moderate-severe TBI assessed with the BTACT at 1 and 2 years post-injury supported its feasibility for inclusion in a large-scale TBI study and provided preliminary evidence that BTACT scores demonstrate expected associations with injury severity and yield more fine-grained assessment of cognition (i.e., a higher ceiling) than the Functional Independence Measure (FIM) Cognition subscale, a clinician-rated assessment of functional cognition in brain injury rehabilitation settings.<sup>12</sup>

The goal of this study was to investigate the validity and utility of the BTACT for detecting cognitive sequelae of TBI. We used data from the Transforming Research and Clinical Knowledge in TBI (TRACK-TBI) study, a multi-center prospective study that enrolled patients ranging in age and TBI severity from U.S. Level 1 trauma centers. At 6 months post-injury, participants completed the BTACT and a battery of in-person neuropsychological tests, enabling investigation of the convergent and discriminant validity of the BTACT across diverse criterion measures. Our specific objectives were to: 1) validate the factor structure of the BTACT to inform the instrument's factorial validity and considerations for how it should be scored; 2) identify the degree to which the factor structure of the BTACT was invariant across groups of interest (i.e., TBI vs. trauma control groups, TBI severity strata) to inform interpretation of group differences in test performance; 3) evaluate the degree of association between scores on the BTACT and other

neuropsychological measures to inform its convergent and discriminant validity; and 4) compare the relative sensitivity of the BTACT and other cognitive tests to TBI and TBI severity, to inform its relative utility in detecting the cognitive sequelae of TBI.

Based on prior published work on the BTACT, we hypothesized that the instrument would manifest a 2-factor structure comprising Episodic Memory and Executive Functioning (EF) factors.<sup>8</sup> However, because of the conceptual appeal of a global BTACT composite and prior psychometric support for a composite, we also considered 1-factor and bifactor models.<sup>13</sup> Additionally, we expected that performance of the BTACT would correlate more strongly with other neurocognitive tests than with self-report or interview-based outcomes, and more specifically that the memory trials of the BTACT would correlate more highly with in-person verbal episodic memory tests than in-person tests of other cognitive domains. Finally, we anticipated that the subscales of the BTACT would manifest variable sensitivity to TBI and TBI severity, with predominantly small effect sizes, given the relatively mild TBI severity of most participants alongside the small cognitive effect sizes expected in the chronic recovery phase.<sup>14,15</sup> Because the episodic memory subtest of the BTACT is briefer than the in-person episodic memory test, we also expected that any group differences would be smaller in magnitude for the BTACT than the in-person memory test.

## Methods

### Participants

The prospective, multi-center TRACK-TBI U01 study enrolled brain injured, orthopedic, and uninjured control participants from 18 Level 1 trauma centers in the United States from February 26, 2014 to July 27, 2018. The institutional review board of each institution approved the study. Inclusion/exclusion criteria are described in detail elsewhere.<sup>14</sup> Participants aged  $\geq 17$  years in the TBI group ( $n=2553$ ) and the orthopedic trauma control group (OTC;  $n=299$ ) were considered for this study. Participants who were eligible for completion of the study's Comprehensive Assessment Battery (i.e., who were living, not yet withdrawn from the study, and sufficiently cognitively intact to participate in neuropsychological testing) were deemed eligible for completion of the BTACT at 6 months post-injury. Of 2852 study participants, 2488 (2217 TBI, 271 OTC) were eligible for BTACT assessment (i.e., living, still in the study, and sufficiently functional to engage in testing); 1592 (1422 TBI, 170 OTC) completed  $\geq 1$  BTACT subtests and were included in our analyses. Supplementary Figure S1 depicts the recruitment and follow-up flow diagram. Table 1 summarizes the characteristics and degree of similarity between the TBI and OTC samples.

### Study protocol

Participants were enrolled within 24 h of injury and performed serial neuropsychological assessments at 2 weeks and 3, 6, and 12 months post-injury. (The 3-month follow-up, conducted by phone, did not assess cognition.) The BTACT was added to the 6-month follow-up protocol to provide preliminary data on its validity and utility. The BTACT was to be performed  $\pm 7$  days of the 6-month in-person assessment, which was achieved in 98.1% of participants with BTACT data (time from in-person to phone assessment median = 2 days;  $M=2.3$ ,  $SD=3.2$ ). All seven BTACT subtests were complete in 99.4% of BTACT assessments.

### Neuropsychological assessment measures

**BTACT.** The cognitive tests included in the current study are listed in Table 2. The BTACT comprises seven subtests, several

TABLE 1. SAMPLE CHARACTERISTICS FOR THE TRAUMATIC BRAIN INJURY (TBI) AND ORTHOPEDIC TRAUMA CONTROL (OTC) GROUPS

	TBI (N=1422) M (SD) or N (%)	OTC (N=170) M (SD) or N (%)	P
Age, years	41.0 (17.2)	40.2 (15.0)	0.907
Sex, male	942 (66%)	108 (64%)	0.494
Race			0.932
White	1099 (77%)	130 (76%)	
Black	234 (16%)	30 (18%)	
Other/unknown	89 (6%)	10 (6%)	
Hispanic	263 (19%)	41 (24%)	0.079
Insurance type			0.280
Insured	917 (66%)	108 (65%)	
Uninsured	290 (21%)	29 (17%)	
Medicare/other	187 (13%)	29 (17%)	
Years of education	13.6 (2.8)	14.0 (2.8)	0.082
Previous TBI	287 (21%)	31 (19%)	0.480
Neurodevelopmental disorder	124 (9%)	16 (9%)	0.774
Psychiatric history	326 (23%)	40 (24%)	0.847
Cause of injury			<0.001
MVC (occupant)	440 (31%)	24 (14%)	
MCC	127 (9%)	22 (13%)	
MVC (cyclist or pedestrian)	230 (16%)	14 (8%)	
Fall	368 (26%)	58 (34%)	
Assault	87 (6%)	2 (1%)	
Other/unknown	170 (12%)	50 (29%)	
TBI severity group			<0.001
GCS 3–8	123 (10%)	0 (0%)	
GCS 9–12	44 (4%)	0 (0%)	
GCS 13–15 CT+	377 (31%)	0 (0%)	
GCS 13–15 CT-	681 (56%)	170 (100%)	
Loss of consciousness	1184 (88%)	0 (0%)	<0.001
Posttraumatic amnesia	1029 (81%)	0 (0%)	<0.001
AIS head/neck			<0.001
Mean (SD)	2.4 (1.4)	0.1 (0.4)	
Median (Q1,Q3)	2 (2, 3)	0 (0, 0)	
ISS Total			<0.001
Mean (SD)	13.7 (9.6)	7.2 (5.9)	
Median (Q1,Q3)	11 (6, 17)	5 (4, 10)	
ISS Peripheral			<0.001
Mean (SD)	5.8 (7.4)	7.0 (5.7)	
Median (Q1,Q3)	4 (1, 9)	5 (4, 10)	
Highest level of care			<0.001
Emergency department	352 (25%)	61 (36%)	
Inpatient unit	552 (39%)	97 (57%)	
Intensive care unit	518 (36%)	12 (7%)	
Injury-related litigation <sup>a</sup>	246 (20%)	25 (17%)	0.442

<sup>a</sup>Collected at 12 months post-injury.

AIS/ISS scores only available on patients admitted to the hospital. Statistical significance by Mann-Whitney and Fisher's exact tests as appropriate. Participants in the OTC group had no CT scans so were assumed GCS 13–15 CT- for the purpose of placement in the table.

AIS, Abbreviated Injury Scale; CT, computed tomography; GCS, Glasgow Coma Scale; ISS, Injury Severity Score; MCC, motorcycle crash; MVC, motor vehicle collision.

TABLE 2. LIST OF COGNITIVE ASSESSMENTS

	Theoretical construct(s) assessed
Phone tests (BTACT) <sup>a</sup>	
Word list recall (RAVLT learning trial 1 and delayed recall)	Verbal episodic memory
Digits span backward (WAIS-III)	Auditory working memory
Category fluency/Animal Naming Test	Semantic verbal fluency/semantic memory retrieval
Number Series	Inductive reasoning
30 Seconds and Counting	Processing speed
In-person paper-and-pencil tests	
RAVLT	Verbal episodic memory
Trail Making Test	Psychomotor processing speed and set-shifting
WAIS-IV Processing Speed Index	Psychomotor processing speed
In-person computerized tests (NIH Toolbox)	
Picture Vocabulary	Receptive vocabulary, estimated general verbal intelligence
Pattern Comparison	Processing speed
Picture Sequence Memory	Visual episodic memory
List Sorting	Working memory
Flanker Inhibitory Control and Attention Test	Attention, inhibitory control
Dimensional Change Card Sorting Test	Cognitive flexibility

<sup>a</sup>The BTACT Stop and Go subtest was administered but was not included in the present analyses because audio recordings were not available to score response times.

BTACT, Brief Test of Adult Cognition by Telephone; NIH, National Institutes of Health; RAVLT, Rey Auditory Verbal Learning Test; WAIS, Wechsler Adult Intelligence Scale.

adapted from widely used paper-and-pencil tests of neuropsychological functioning. Form A of the BTACT was administered. The BTACT subtests measure verbal episodic memory (via one immediate recall trial of the Rey Auditory Verbal Learning Test [RAVLT] Form 1 and one delayed recall trial, score range 0–15 for each trial), auditory working memory (i.e., backward trials from the Wechsler Adult Intelligence Scale–Third Edition [WAIS-III] Digit Span test, scored as longest correct span from 2–8), semantic verbal fluency (Animal Naming Test), inductive reasoning (Number Series, in which the participant completes the pattern in a series of numbers, range 0–5), and processing speed (30 Seconds And Counting Task, or 30-SACT; the number of digits produced by counting backward from 100 in 30 sec). The optional Stop and Go Switch Task (SGST) was administered but not included in analyses because audio recordings needed to score response latencies were not collected.

**Other cognitive performance measures.** Participants completed a series of neuropsychological tests at the in-person visit, including paper-and-pencil tests of verbal episodic memory (RAVLT<sup>16</sup>), processing speed (Wechsler Adult Intelligence Scale–Fourth Edition Processing Speed Index [WAIS-IV PSI]),<sup>17</sup> and executive functioning (Trail Making Test [TMT]).<sup>18</sup> Different RAVLT forms were administered across the 2-week (Form 2),<sup>19</sup> 6-month in-person (Form 3), and 6-month BTACT (Form 1)

assessment. Prior studies have found these alternate forms to be equivalent and serial administration of alternate forms to cause minimal practice effects.<sup>19,20</sup> Additionally, participants completed modules of the computerized National Institutes of Health (NIH) Toolbox Cognition Battery intended to estimate receptive vocabulary/estimated verbal intellectual functioning (Picture Vocabulary Test, TPVT), processing speed (Pattern Comparison), visual episodic memory (Picture Sequence Memory Test [PSMT]), working memory (List Sorting), attention/inhibitory control (Flanker Inhibitory Control and Attention Test [Flanker]), and cognitive flexibility (Dimensional Change Card Sort Test [DCCS]).<sup>21</sup> All subtests except for the TPVT were aggregated into a Fluid Reasoning Composite score by the NIH Toolbox scoring software. Unstandardized scale scores (i.e., standardized against the entire NIH Toolbox normative sample but not demographically corrected; normative sample  $M=100$  and  $SD=15$ ) were used for all NIH Toolbox indices in this article.

**Symptoms.** Self-report symptom scales were administered at the in-person visit to estimate the severity of symptoms in the domains of TBI/injury symptoms (Rivermead Post Concussion Symptoms Questionnaire [RPQ]; ratings of "1" were not counted in total scores),<sup>22</sup> emotional distress (18-item Brief Symptom Inventory [BSI-18]),<sup>23</sup> depression (9-item Patient Health Questionnaire [PHQ-9]),<sup>24,25</sup> post-traumatic stress disorder (PTSD Checklist for DSM-5 [PCL-5]),<sup>26</sup> TBI-related quality of life (6-item Quality of Life after Brain Injury Overall Scale [QOLIBRI-OS]),<sup>27</sup> health-related quality of life (12-item Short Form Health Survey [SF-12]),<sup>28</sup> and general life satisfaction (Satisfaction With Life Scale [SWLS]).<sup>29</sup>

**Functional outcome.** The Glasgow Outcome Scale-Extended (GOS-E) was used to evaluate the impact of the traumatic injury on daily functioning.<sup>30</sup> The GOS-E was assessed via structured interview with patients or proxies.<sup>31</sup> Respondents were asked to report new injury-related dependence or difficulties in several domains of life function. For the current study, changes in function were counted irrespective of whether they resulted from TBI or other injury-related factors (e.g., peripheral injuries). Score ranges in this sample were 3–8, where 3=Lower Severe Disability and 8=Upper Good Recovery.

### Statistical analysis

All available BTACT data were used in factor analyses, measurement invariance analyses, and group comparisons. The 1.9% of case patients who completed the BTACT more than  $\pm 7$  days of the in-person neuropsychological assessment were excluded from analyses comparing BTACT with other clinical outcomes.

**Factor structure and group invariance of the BTACT.** We performed confirmatory factor analysis (CFA) to cross-validate the fit of the 2-factor structure of the BTACT previously derived by the instrument's authors and validated in a general community sample.<sup>8,32</sup> This model posits the existence of two correlated factors, an Episodic Memory factor on which both the immediate and delayed recall trial of the RAVLT load, and an EF factor on which all other core subtests load. Because of the conceptual appeal of a global BTACT composite score, as well as precedence in the literature for computing a single composite score,<sup>7,13</sup> we also considered the possibility of a 1-factor model or a bifactor model (on which all items load on a general factor as well as independent specific factors to account for residual correlations between specified items). Factor models were tested using maximum likelihood estimation with robust standard errors (MLR). Evidence of good fit was determined *a priori* to be established from commonly recommended fit statistics: root mean square error of approximation (RMSEA)

<0.06 and Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) >0.95.<sup>33</sup> We anticipated that the  $\chi^2$  Goodness of Fit Test would be significant (i.e., suggestive of inadequate model fit) as an artifact of this index's oversensitivity to minor deviations in fit in large samples.

Measurement invariance modeling was carried out to establish the degree to which the BTACT reflects the same constructs between different TBI severity groups and the OTC group. This was tested through a series of CFA models in which we imposed increasingly stringent constraints on the equivalence of model parameters between groups: 1) that the same items load on the same factors, allowing other parameters to vary between groups (configural invariance); 2) that, additionally, factor loadings are equivalent between groups (weak invariance); 3) that, additionally, item intercepts are equivalent between groups (strong invariance); and 4) that, additionally, residual variances are equivalent between groups (strict invariance). Strong or strict invariance is generally considered to be a good outcome because it allows one to interpret differences between groups in factor means to fully reflect differences between groups on the underlying constructs of interest. By convention, we *a priori* decided to conclude that the BTACT had a higher level of invariance if imposing the additional constraints of that invariance model did not decrease model fit by more than 0.01 in CFI<sup>34</sup> or 0.015 in RMSEA.<sup>35</sup> Factor and group invariance modeling was conducted using *Mplus* version 8.3.<sup>36</sup>

The sample was stratified into four groups: the OTC group and three TBI strata. TBI strata were defined using historical markers of injury severity: admission Glasgow Coma Scale (GCS) score of 13–15 without acute intracranial abnormalities on computed tomography (CT) scans (mild TBI [mTBI] CT-), GCS 13–15 with acute intracranial abnormalities (mTBI CT+), and GCS 3–12 (moderate-severe TBI). Differences between groups in latent mean episodic memory and executive function performances were compared in *Mplus* for the strict invariance model.

**Convergent and discriminant validity of the BTACT.** To evaluate to what degree the BTACT measures what it is purported to measure (i.e., CoI) we examined correlations between BTACT subtest and factor scores and other available neuropsychological measures collected at the 6-month follow-up (e.g., self-report, interview, and cognitive performance measures). Pearson correlations ( $r$ ) are reported for all measures except for the ordinal GOS-E, for which we computed Spearman correlations ( $\rho$ ). Given the study aims, we were primarily interested in the overall and relative magnitude of associations.

**Relationship between BTACT performance and injury (TBI, orthopedic trauma) and TBI severity.** Finally, we evaluated the sensitivity of the BTACT to TBI severity (i.e., PoU) by comparing performance among TBI severity groups and between each TBI severity group and the OTC group. The sensitivity of the BTACT was compared with other available neurocognitive measures. Group comparisons were carried out using general linear models (GLM) adjusted for age, sex, and education and with patterns of missing outcomes adjusted using propensity weights.<sup>37</sup> Propensity weights were derived separately by measure type (GOS-E, self-report, neuropsychological) from boosted logistic regression models predicting completion versus non-completion of outcomes from enrollment site and demographic, history, and injury variables. Weights were proportional to the inverse of the probability of measure completion and normed so the sum equaled the number of cases with the measure actually completed.

Group effect sizes were reported as Cohen's  $d$  (with 95% confidence intervals [CIs]). Comparisons that remained significant after multiple comparison correction (within outcome assessment type and time-point using a 5% false discovery rate)<sup>38</sup> are bolded in the tables and discussed.

Propensity probabilities were generated using the “twang” software package developed for R (version 3.2.2), which was accessed via script files generated by a SAS macro package (version 9.3). Other statistical analyses were conducted in IBM SPSS Statistics for Windows (version 19).

**Results**

*Factor and group invariance modeling*

CFA of the BTACT indicated good fit of the 2-factor model, RMSEA=0.056 (90% CI=0.043–0.071), CFI=0.982, TLI=0.970. A traditional 1-factor model (where residual variances are uncorrelated) fit poorly, RMSEA=0.238, CFI=0.687, TLI=0.463. A 3-factor bifactor model with a general factor (on which all items loaded) and two independent specific factors (where memory items loaded onto one specific factor, and EF items on a second specific factor) was uninterpretable, with 3 of 4 loadings on the EF factor trivial in size (<0.20) and in opposing directions. Thus, we settled on a 1-factor model that allowed for correlated residual variances between the Immediate and Delayed Recall trials (i.e., the equivalent of a General factor from a bifactor model with 1 specific Memory factor), which fit comparably to the 2-factor model, RMSEA=0.048 (90% CI 0.033–0.064), CFI=0.988, TLI=0.987. The Chi-square test of model fit was significant in all models tested ( $p < 0.001$ ), as expected in this large sample. The 2-factor and 1-factor (with correlated residuals) models are depicted in Figure 1. In the 2-factor model, the Episodic Memory and EF factors were correlated 0.52, with all loadings significant and ranging from 0.58 (Animal Naming Test) to 0.92 (RAVLT Immediate Recall). In the 1-factor model, loadings on the General factor ranged from 0.36 to 0.71. For reference, Table 3 depicts the inter-correlations among BTACT subset and factor scores.

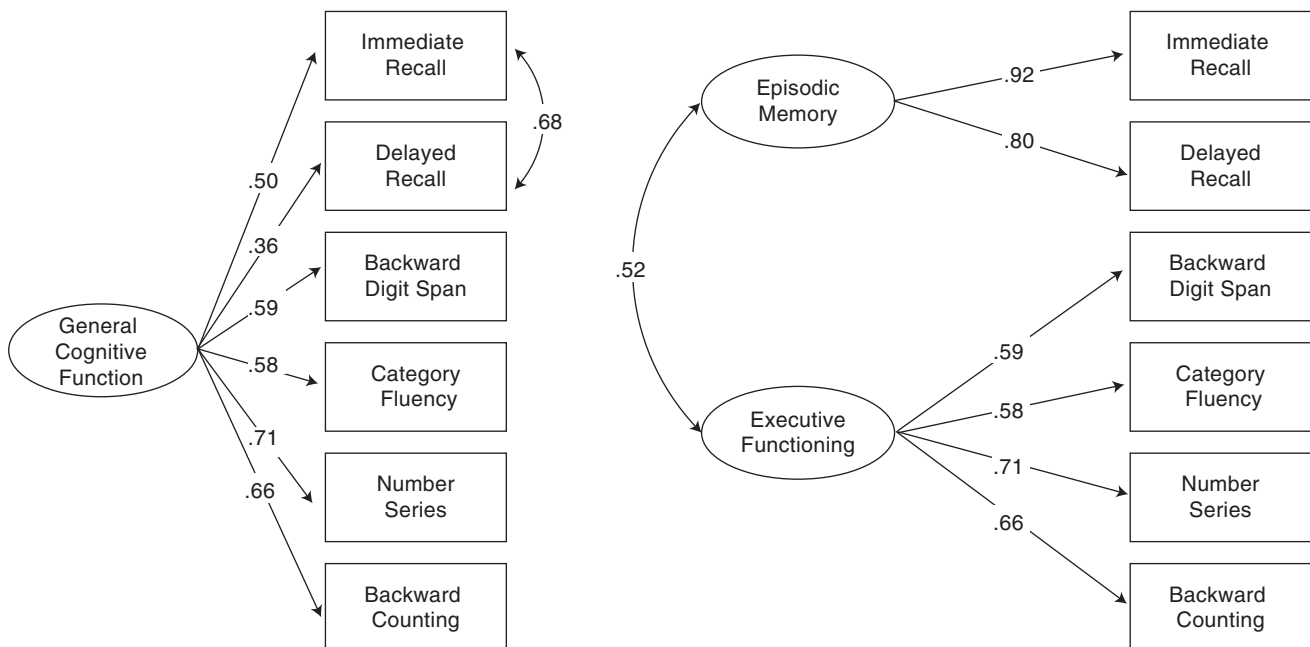
Table 4 provides fit statistics for the measurement invariance analysis of both factor models by group (mTBI CT-, mTBI CT+, moderate-severe TBI, and OTC). Invariance analyses supported conclusions of strict invariance across groups for both the 2-factor and 1-factor model. Model-estimated latent mean differences

comparing the groups were evaluated, which reflect factor *SD* units referenced to each target group ( $M \text{ difference} = M_{\text{diff}}$ ) as described below. The latent mean of all three factors were indicative of significantly poorer performance for moderate-severe TBI versus the mTBI CT- group ( $M_{\text{diff}} = -0.31 - -0.37$ ) and the OTC group ( $M_{\text{diff}} = -0.48 - -0.59$ ), whereas the moderate-severe TBI group only performed significantly worse than the mTBI CT+ group in Episodic Memory ( $M_{\text{diff}} = -0.29$ ). The mTBI CT- and mTBI CT+ groups were not significantly different in any latent factor mean ( $p \geq 0.082$ ), but both mTBI groups performed significantly worse than OTCs in General Cognitive Function (mTBI CT-  $M_{\text{diff}} = -0.25, p < 0.001$ ; mTBI CT+  $M_{\text{diff}} = -0.35, p = 0.001$ ) and EF (mTBI CT-  $M_{\text{diff}} = -0.26, p = 0.012$ ; mTBI CT+  $M_{\text{diff}} = -0.36, p = 0.001$ ). Factor scores were estimated in *Mplus* from the strict invariance model for further analysis.

*Convergent and discriminant validity of the BTACT*

Correlational analyses established the degree to which the BTACT factor and subtest scores correlated with self/informant-report measures (Table 5), paper-and-pencil neurocognitive tests (Table 6), and computerized neurocognitive tests (NIH Toolbox; Table 7) collected within 7 days of the BTACT. As predicted, correlations between indices on the BTACT and other cognitive tests were generally stronger than BTACT/self-report associations ( $M \text{ BTACT/self-informant } r = 0.18, M \text{ BTACT/paper-and-pencil cognitive } r = 0.41, M \text{ BTACT/computerized cognitive } r = 0.37$ ).

Because the most closely aligned phone and in-person assessments were those of verbal episodic memory, we first examined the relationship between BTACT RAVLT and in-person RAVLT assessments, as compared with BTACT RAVLT and in-person tests of other cognitive domains. Consistent with expectation, BTACT RAVLT List 1 immediate recall was most robustly related to in-person RAVLT scores ( $r = 0.46$  for Trial 1 recall). That the loading of this subscale onto the BTACT Memory factor was nearly perfect, and external correlates of the BTACT Memory factor score similar to BTACT RAVLT Trial 1 recall score, implies that the BTACT



**FIG. 1.** The 1-factor (with correlated residuals) and 2-factor models of the Brief Test of Adult Cognition by Telephone (BTACT).

TABLE 3. CORRELATION<sup>a</sup> BETWEEN BTACT FACTOR AND SUBSCALE SCORES

	A	B	C	D	E	F	G	H
A. General factor								
B. Episodic Memory factor	0.64							
C. RAVLT Trial 1 Recall	0.59	0.98						
D. RAVLT Delayed Recall	0.44	0.85	0.73					
E. Executive Functioning factor	>0.99	0.63	0.58	0.46				
F. Digits Backward	0.67	0.39	0.35	0.27	0.68			
G. Categorical Fluency	0.68	0.36	0.32	0.23	0.68	0.30		
H. Number Series	0.80	0.38	0.33	0.24	0.80	0.43	0.39	
I. 30 Seconds and Counting	0.76	0.36	0.29	0.19	0.76	0.36	0.41	0.48

<sup>a</sup>Cell sample size range 1363–1592. All correlations were significant with  $p < 0.001$ . BTACT, Brief Test of Adult Cognition by Telephone; RAVLT, Rey Auditory Verbal Learning Test.

Memory factor can be considered relatively equivalent to the Trial 1 recall score. BTACT RAVLT Delayed Recall Trial score also correlated most strongly with in-person RAVLT Delayed Recall scores ( $r = 0.42$ , vs.  $M r$  with other cognitive domains = 0.19). Overall, our hypotheses were supported, as the BTACT memory indices correlated more strongly with in-person memory tests ( $M r = 0.48$ ) than in-person tests of other domains ( $M r = 0.25$ ).

Other BTACT subscale scores (Digits Backward, Category Fluency, Number Series, and 30-SACT) correlated similarly with in-person tests of a variety of domains ( $M r$  with in-person memory tests = 0.36,  $M r$  with other in-person tests = 0.38). The BTACT EF factor, comprising these four subtests, correlated even more strongly with tests of a variety of domains, including processing speed (WAIS-IV PSI  $r = 0.60$ ), verbal episodic memory ( $r = 0.51$ – $0.60$ ,  $M = 0.51$ ), fluid reasoning/executive functioning ( $r = 0.40$ – $0.67$ ,  $M = 0.52$ ), and receptive vocabulary ( $r = 0.53$ ).

*Sensitivity of the BTACT to TBI and TBI severity*

Finally, we investigated the relative sensitivity of the BTACT versus other cognitive tests to detect TBI sequelae and severity. Table 8 provides descriptive statistics in cognitive performance across phone (BTACT), paper-and-pencil, and computerized cognitive tests. Table 9 depicts group comparisons (Cohen’s  $d$  and 95% CIs) among the three TBI severity groups and between each TBI

subgroup and the OTC group. mTBI (GCS 13–15) CT-, and CT+ groups were only significantly different on one test, with the CT- group performing worse on the NIH Toolbox Dimensional Change Card Sorting test ( $d = 0.33$ ). Across modes of assessment, differences between both mTBI groups and the moderate-severe TBI (GCS 3–12) group were similar in direction and pattern of significance, although the magnitude of group differences tended to be largest for the paper-and-pencil tests ( $M d = 0.54$ ) and smaller for the BTACT indices ( $M d = 0.27$ ) and computerized tests ( $M d = 0.30$ ).

Regarding TBI versus OTC group comparisons, the BTACT achieved the most consistent and strongest group differences in its General and EF factors, with poorer performance for the mTBI CT- ( $d = -0.25$  –  $-0.26$ ), mTBI CT+ ( $d = -0.21$  –  $-0.32$ ), and moderate-severe TBI groups ( $d = -0.66$  –  $-0.69$ ) as compared with OTCs. Most BTACT indices reflected significantly poorer performance for the moderate-severe TBI (vs. the OTC) group, although effect sizes for moderate-severe TBI versus OTC groups were stronger for paper-and-pencil tests ( $M d = -0.62$ ) as compared with the BTACT ( $M d = -0.43$ ) and computerized tests ( $M d = -0.33$ ). The higher sensitivity of paper-and-pencil tests to moderate-severe TBI was true even for the comparable in-person episodic memory (RAVLT  $M d = -0.65$ ) and BTACT memory indices ( $M d = -0.36$ ).

Not surprisingly, differences in performance between mTBI and OTC groups were weaker and more modest than those involving

TABLE 4. FIT STATISTICS FOR MEASUREMENT INVARIANCE OF THE BTACT FACTOR MODELS ACROSS TRAUMATIC BRAIN INJURY (TBI) AND ORTHOPEDIC TRAUMA CONTROL (OTC) GROUPS

	-2LL	k	BIC	RMSEA	ΔRMSEA	CFI	ΔCFI	TLI
1-factor model <sup>a</sup>								
Configural	-19976	76	40501	0.036		0.993		0.988
Weak	-19983	61	40406	0.027	-0.009	0.995	0.002	0.993
Strong	-19997	46	40326	0.034	0.007	0.989	-0.006	0.989
Strict	-20017	25	40215	0.038	0.004	0.981	-0.008	0.986
2-factor model								
Configural	-19984	72	40488	0.046		0.988		0.980
Weak	-19988	63	40431	0.040	-0.006	0.989	0.001	0.985
Strong	-19998	51	40364	0.041	0.001	0.985	-0.004	0.984
Strict	-20016	33	40270	0.043	0.002	.0978	-0.007	0.982

<sup>a</sup>1-factor model with correlated residuals between the two episodic memory trials.

Four groups were included in the model: the OTC group and three TBI groups, defined by Glasgow Coma Scale (GCS) score 13–15 without acute intracranial findings on computed tomography (CT) scans, GCS 13–15 with acute intracranial findings on CT, and GCS 3–12.

BIC, Bayesian information criterion; BTACT, Brief Test of Adult Cognition by Telephone; CFI, Comparative Fit Index; k, free parameters; LL, log likelihood; RMSEA, root mean square error of approximation; TLI, Tucker-Lewis index.

TABLE 5. CORRELATION<sup>a</sup> BETWEEN BTACT FACTOR AND SELF-REPORT/INFORMANT REPORT OUTCOMES

<i>BTACT factor/subtest</i>	<i>BSI-18 GSI</i>	<i>RPQ</i>	<i>PHQ-9</i>	<i>QoLIBRI-OS</i>	<i>PCL-5</i>	<i>SF-12 PCS</i>	<i>SF-12 MCS</i>	<i>GOS-E<sup>b</sup></i>
General factor	-0.21	-0.30	-0.23	0.25	-0.25	0.26	0.14	0.25
Episodic Memory factor	-0.18	-0.22	-0.19	0.22	-0.21	0.18	0.14	0.23
RAVLT Trial 1 Recall	-0.17	-0.21	-0.18	0.20	-0.20	0.16	0.13	0.21
RAVLT Delayed Recall	-0.11	-0.13	-0.13	0.15	-0.13	0.12	0.09	0.17
Executive Functioning factor	-0.21	-0.30	-0.23	0.25	-0.25	0.26	0.14	0.25
Digits Backward	-0.16	-0.20	-0.18	0.18	-0.17	0.14	0.12	0.16
Categorical Fluency	-0.13	-0.21	-0.15	0.16	-0.14	0.23	0.06	0.17
Number Series	-0.20	-0.26	-0.20	0.20	-0.23	0.21	0.12	0.20
30 Seconds and Counting	-0.16	-0.22	-0.17	0.18	-0.18	0.22	0.10	0.19

<sup>a</sup>Cell sample size range 985–1410. Assuming smallest cell sample size, correlations above 0.082 are significant at 0.01; correlations above 0.105 are significant at 0.001.

<sup>b</sup>Spearman’s rho; all other values are Pearson *r*.

BSI-18 GSI, 18-item Brief Symptom Inventory Global Severity Index; GOS-E, Glasgow Outcome Scale–Extended; PCS, Physical Component Score; PCL-5, Post-traumatic Stress Disorder Checklist for DSM-5; PHQ-9, Patient Health Questionnaire depression scale; QoLIBRI-OS, Quality of Life after Brain Injury Overall Scale; RAVLT, Rey Auditory Verbal Learning Test; RPQ, Rivermead Post Concussion Symptom Questionnaire; SF-12, 12-item Short Form Health Survey version 2.

moderate-severe TBI. BTACT and paper-and-pencil modes of assessment each produced at least one measure with significant and comparable sensitivity to mTBI, with the most robust associations between both mTBI and the OTC group for the BTACT General and EF factors ( $d = -0.24$  to  $-0.32$ ) and paper-and-pencil TMT Part A ( $d = 0.23$  to  $0.31$ ). On the NIH Toolbox, two mTBI versus OTC comparisons were in the opposite direction of expectation, whereby the mTBI CT- group performed significantly better than the OTC group in deductive reasoning (DCCS  $d = 0.34$ ) and the mTBI CT+ group performed significantly better than the OTC group in response inhibition (Flanker  $d = 0.26$ ).

**Discussion**

We leveraged a large Level 1 trauma center patient sample from the TRACK-TBI study to examine the construct validity of the BTACT and its utility for detecting cognitive sequelae 6 months post-TBI. Confirmatory factor analysis supported both a 2-factor model previously established in a general community sample<sup>8,9</sup> and a 1-factor model supported by other work.<sup>13</sup> This indicates that

performance on the BTACT subtests can be conceptualized as reflecting either two moderately correlated dimensions, historically labeled Episodic Memory (comprising an abbreviated RAVLT word list memory test) and Executive Functioning (EF), or as reflecting a single common dimension (which we labeled General Cognitive Function). The findings support scoring the BTACT based on either model, depending on one’s theoretical views and practical goals.

Analyses of the convergent and discriminant validity of the BTACT factors and subtests support conclusions that the memory scores appear to index verbal episodic memory. This was supported by findings of substantially stronger correlations between BTACT and in-person RAVLT performances than with other cognitive performance or self-report measures. However, the magnitude of correlations between BTACT and in-person RAVLT test trials was modest ( $r = 0.32$ – $0.51$ ), perhaps because the abbreviated nature of the BTACT memory test restricts the score range and, consequently, potential for BTACT memory scores to correlate with other measures. That other investigators have found smaller alternate forms reliability coefficients between Trial 1 RAVLT

TABLE 6. CORRELATION<sup>a</sup> BETWEEN BTACT FACTOR AND SUBSCALE SCORES AND PAPER-AND-PENCIL NEUROCOGNITIVE TEST PERFORMANCE

<i>BTACT factor/subtest</i>	<i>Trail Making Test</i>		<i>WAIS-IV PSI</i>	<i>Rey Auditory Verbal Learning Test (RAVLT) Recall</i>								
	<i>Part A time</i>	<i>Part B time</i>		<i>Trial 1</i>	<i>Trial 2</i>	<i>Trial 3</i>	<i>Trial 4</i>	<i>Trial 5</i>	<i>Trial 1-5 sum</i>	<i>List B</i>	<i>Short delay</i>	<i>Long delay</i>
General factor	-0.51	-0.57	0.60	0.52	0.55	0.54	0.53	0.53	0.60	0.52	.52	.51
Episodic Memory factor	-0.27	-0.31	0.33	0.49	0.52	0.50	0.49	0.47	0.55	0.45	.47	.48
RAVLT Trial 1 Recall	-0.23	-0.27	0.27	0.46	0.48	0.45	0.44	0.42	0.51	0.41	.41	.42
RAVLT Delayed Recall	-0.15	-0.19	0.20	0.32	0.38	0.38	0.38	0.37	0.42	0.33	.40	.42
Executive Functioning factor	-0.50	-0.57	0.60	0.51	0.55	0.54	0.53	0.53	0.60	0.52	.52	.51
Digits Backward	-0.31	-0.35	0.34	0.29	0.33	0.32	0.33	0.33	0.36	0.32	.29	.30
Categorical Fluency	-0.36	-0.39	0.42	0.37	0.40	0.41	0.40	0.39	0.45	0.37	.42	.41
Number Series	-0.37	-0.44	0.46	0.34	0.36	0.36	0.36	0.35	0.40	0.37	.36	.34
30 Seconds and Counting	-0.43	-0.48	0.48	0.33	0.34	0.34	0.33	0.34	0.38	0.32	.33	.31

<sup>a</sup>Cell sample size range 1150–1361. Assuming smallest cell sample size, correlations above 0.076 are significant at 0.01; correlations above 0.097 are significant at 0.001.

BTACT, Brief Test of Adult Cognition by Telephone; PSI, Processing Speed Index; RAVLT, Rey Auditory Verbal Learning Test; WAIS-IV, Wechsler Adult Intelligence Scale-Fourth Edition.



TABLE 7. CORRELATION<sup>a</sup> BETWEEN BTACT FACTOR AND SUBSCALE SCORES AND NIH TOOLBOX COGNITIVE<sup>b</sup> SCALES

<i>BTACT factor/subtest</i>	<i>Fluid Reasoning Composite<sup>c</sup></i>	<i>Dimensional Change Card Sort</i>	<i>Flanker</i>	<i>List Sorting</i>	<i>Pattern Completion</i>	<i>Picture Sequence Memory</i>	<i>Picture Vocabulary</i>
General factor	0.67	0.51	0.41	0.59	0.40	0.51	0.53
Episodic Memory factor	0.41	0.29	0.24	0.35	0.19	0.33	0.35
RAVLT Trial 1 Recall	0.36	0.24	0.20	0.31	0.14	0.29	0.33
RAVLT Delayed Recall	0.28	0.17	0.11	0.20	0.12	0.27	0.23
Executive Functioning factor	0.67	0.51	0.41	0.59	0.40	0.51	0.53
Digits Backward	0.38	0.26	0.21	0.38	0.20	0.29	0.35
Categorical Fluency	0.43	0.31	0.20	0.40	0.25	0.34	0.41
Number Series	0.54	0.39	0.31	0.48	0.28	0.42	0.47
30 Seconds and Counting	0.53	0.46	0.41	0.43	0.38	0.36	0.31

<sup>a</sup>Cell sample size range 702–891. Assuming smallest cell sample size, correlations above 0.097 are significant at 0.01; correlations above 0.124 are significant at 0.001.

<sup>b</sup>Unstandardized scale scores.

<sup>c</sup>Fluid Reasoning Composite aggregates the Dimensional Change Card Sort, Flanker, List Sorting, Pattern Comparison, and Picture Sequence Memory tests.

BTACT, Brief Test of Adult Cognition by Telephone; NIH, National Institutes of Health; RAVLT, Rey Auditory Verbal Learning Test.

( $r=0.31$ ) than Trial 1–5 total score ( $r=0.78$ ) supports the hypothesis that restricted score range/reliability diminished the BTACT versus in-person RAVLT correlations in this study.<sup>20</sup> However, we cannot rule out the possibility that the BTACT memory test taps more into verbal attention span (vs. learning processes) given its

administration of only one learning trial,<sup>39,40</sup> as compared with the full-length RAVLT, which offers an opportunity to assess learning from repetition.

Perhaps also related to the limited score range of the BTACT RAVLT or more limited focus on learning and long-term memory

TABLE 8. DESCRIPTIVE STATISTICS FOR COGNITIVE PERFORMANCE AT 6 MONTHS POST-INJURY BY INJURY GROUP

	<i>mTBI (GCS 13–15) CT-</i>			<i>mTBI (GCS 13–15) CT+</i>			<i>Moderate-Severe TBI (GCS 3–12)</i>			<i>OTC</i>		
	N	M	SD	N	M	SD	N	M	SD	N	M	SD
<b>BTACT</b>												
General factor	681	-0.29	0.86	353	-0.46	0.94	167	-0.68	0.86	170	-0.07	0.86
Episodic Memory factor	681	-0.13	0.91	353	-0.22	0.95	167	-0.49	0.89	170	-0.04	0.93
RAVLT Trial 1 Recall	681	6.35	2.27	351	6.16	2.35	167	5.60	2.34	170	6.52	2.24
RAVLT Delayed Recall	679	3.51	2.55	347	3.34	2.57	167	2.72	2.34	170	3.73	2.74
Executive Functioning factor	681	-0.31	0.87	353	-0.47	0.96	167	-0.65	0.86	170	-0.07	0.86
Digit Span Backward	681	4.96	1.53	353	4.87	1.64	167	4.77	1.45	170	5.14	1.63
Category Fluency	681	19.5	6.0	353	18.9	6.2	166	17.6	5.6	170	20.8	5.3
Number Series	679	2.54	1.72	345	2.46	1.77	167	2.28	1.65	170	2.84	1.69
30-SACT	679	39.1	11.2	349	37.1	11.8	167	35.4	10.7	170	40.0	10.8
<b>Paper-and-Pencil</b>												
TMT A Raw	661	28.3	13.5	329	32.7	18.0	163	35.6	17.2	168	25.5	9.8
TMT B Raw	658	76.8	44.3	327	86.5	59.0	153	96.6	67.2	168	70.5	42.9
WAIS PSI	659	101.4	16.4	326	98.2	17.3	159	89.7	17.0	167	101.7	14.7
RAVLT Trial 1-5 Sum	670	46.2	11.1	334	43.2	11.5	161	39.2	10.6	169	47.2	12.0
RAVLT Delay Recall	670	9.08	3.53	334	8.29	3.60	161	6.87	3.63	168	9.48	3.70
<b>NIH Toolbox</b>												
Fluid Cognition Composite	424	109.9	13.2	175	107.3	14.0	82	104.6	13.1	108	108.3	12.8
Dimen. Change Card Sort	438	108.7	13.3	184	103.0	12.9	84	104.7	13.9	117	104.6	12.3
Flanker	439	105.9	12.1	189	104.0	13.2	89	99.8	12.1	117	102.9	11.0
List Sorting	440	108.9	12.0	184	107.1	13.8	86	106.8	12.4	117	111.4	13.2
Pattern Comparison	441	105.8	18.9	184	98.3	20.5	84	96.1	19.4	116	104.7	18.4
Picture Sequence Memory	427	104.1	15.3	177	101.8	15.1	82	98.8	15.5	109	103.2	15.6
Picture Vocabulary	440	115.7	11.5	188	116.7	12.7	89	114.9	11.4	118	119.3	12.7

30-SACT, 30 Seconds and Counting test of backward counting; BTACT, Brief Test of Adult Cognition by Telephone; CT-, computed tomography scan negative for acute intracranial findings; CT+, computed tomography scan positive for acute intracranial findings; GCS, admission Glasgow Coma Scale score; mTBI, mild TBI (GCS score 13–15); NIH Toolbox, National Institutes of Health Toolbox Cognition Battery; OTC, orthopedic trauma control; TBI, traumatic brain injury; TMT, Trail Making Test; WAIS-IV PSI, Wechsler Adult Intelligence Scale-Fourth Edition Processing Speed Index.

TABLE 9. COHEN'S *D* EFFECT SIZES (95% CONFIDENCE INTERVALS) OF GROUP COMPARISONS ON COGNITIVE PERFORMANCE

	<i>Comparisons between TBI severity groups</i>			<i>Comparisons between each TBI severity group and the OTC group</i>		
	<i>GCS 13–15: CT- vs. CT+ d<sup>a</sup></i>	<i>GCS 13–15 CT- vs. GCS 3–12 d</i>	<i>GCS 13–15 CT+ vs. GCS 3–12 d</i>	<i>TBI GCS 13–15 CT- vs. OTC d</i>	<i>TBI GCS 13–15 CT+ vs. OTC d</i>	<i>TBI GCS 3–12 vs. OTC d</i>
<b>BTACT</b>						
General factor	0.08 (-0.06, 0.21)	<b>0.43</b> <b>(0.26, 0.61)</b>	<b>0.38</b> <b>(0.18, 0.58)</b>	<b>-0.25</b> <b>(-0.42, -0.08)</b>	<b>-0.31</b> <b>(-0.50, -0.12)</b>	<b>-0.69</b> <b>(-0.92, -0.47)</b>
Episodic Memory factor	0.03 (-0.11, 0.16)	<b>0.33</b> <b>(0.15, 0.50)</b>	<b>0.30</b> <b>(0.10, 0.50)</b>	-0.07 (-0.24, 0.10)	-0.10 (-0.29, 0.09)	<b>-0.42</b> <b>(-0.64, -0.19)</b>
RAVLT Trial 1 Recall	0.03 (-0.11, 0.16)	<b>0.24</b> <b>(0.07, 0.42)</b>	<b>0.22</b> <b>(0.02, 0.42)</b>	-0.05 (-0.22, 0.12)	-0.08 (-0.27, 0.11)	<b>-0.32</b> <b>(-0.54, -0.10)</b>
RAVLT Delayed Recall	0.00 (-0.14, 0.13)	<b>0.27</b> <b>(0.09, 0.45)</b>	<b>0.28</b> <b>(0.09, 0.48)</b>	-0.08 (-0.25, 0.09)	-0.06 (-0.25, 0.13)	<b>-0.34</b> <b>(-0.56, -0.12)</b>
Executive Functioning Factor	0.08 (-0.06, 0.21)	<b>0.38</b> <b>(0.20, 0.55)</b>	<b>0.33</b> <b>(0.13, 0.52)</b>	<b>-0.26</b> <b>(-0.43, -0.09)</b>	<b>-0.32</b> <b>(-0.51, -0.13)</b>	<b>-0.66</b> <b>(-0.88, -0.43)</b>
Digit Span Backward	-0.01 (-0.14, 0.13)	0.03 (-0.14, 0.21)	0.09 (-0.11, 0.29)	-0.09 (-0.26, 0.08)	-0.06 (-0.24, 0.13)	-0.14 (-0.36, 0.08)
Category Fluency	0.08 (-0.05, 0.22)	<b>0.35</b> <b>(0.18, 0.53)</b>	<b>0.27</b> <b>(0.08, 0.47)</b>	-0.17 (-0.34, 0.00)	<b>-0.27</b> <b>(-0.45, -0.08)</b>	<b>-0.59</b> <b>(-0.81, -0.37)</b>
Number Series	-0.07 (-0.21, 0.06)	0.11 (-0.06, 0.29)	<b>0.22</b> <b>(0.02, 0.41)</b>	-0.14 (-0.31, 0.02)	-0.07 (-0.26, 0.12)	<b>-0.27</b> <b>(-0.49, -0.05)</b>
30-SACT	0.09 (-0.05, 0.22)	<b>0.34</b> <b>(0.16, 0.52)</b>	<b>0.25</b> <b>(0.05, 0.45)</b>	-0.08 (-0.25, 0.09)	-0.16 (-0.35, 0.03)	<b>-0.44</b> <b>(-0.67, -0.22)</b>
<b>Paper-and-Pencil</b>						
TMT A Raw	-0.15 (-0.29, -0.01)	<b>-0.52</b> <b>(-0.70, -0.34)</b>	<b>-0.31</b> <b>(-0.51, -0.10)</b>	<b>0.23</b> <b>(0.06, 0.40)</b>	<b>0.31</b> <b>(0.11, 0.50)</b>	<b>0.68</b> <b>(0.45, 0.91)</b>
TMT B Raw	-0.09 (-0.23, 0.05)	<b>-0.47</b> <b>(-0.66, -0.29)</b>	<b>-0.34</b> <b>(-0.54, -0.13)</b>	0.11 (-0.06, 0.28)	0.15 (-0.05, 0.34)	<b>0.46</b> <b>(0.23, 0.69)</b>
WAIS PSI	0.13 (-0.01, 0.27)	<b>0.65</b> <b>(0.47, 0.83)</b>	<b>0.50</b> <b>(0.29, 0.70)</b>	0.02 (-0.15, 0.19)	-0.08 (-0.27, 0.12)	<b>-0.64</b> <b>(-0.87, -0.41)</b>
RAVLT Trial 1-5 Sum	0.09 (-0.05, 0.23)	<b>0.66</b> <b>(0.48, 0.85)</b>	<b>0.60</b> <b>(0.39, 0.81)</b>	-0.05 (-0.22, 0.13)	-0.15 (-0.34, 0.04)	<b>-0.63</b> <b>(-0.86, -0.40)</b>
RAVLT Delay Recall	0.02 (-0.12, 0.16)	<b>0.67</b> <b>(0.49, 0.86)</b>	<b>0.63</b> <b>(0.42, 0.84)</b>	-0.11 (-0.28, 0.07)	-0.15 (-0.34, 0.05)	<b>-0.67</b> <b>(-0.90, -0.44)</b>
<b>NIH Toolbox</b>						
Fluid Cognition Composite	0.12 (-0.07, 0.30)	<b>0.41</b> <b>(0.17, 0.65)</b>	0.28 (0.01, 0.56)	0.10 (-0.11, 0.32)	0.00 (-0.25, 0.25)	-0.29 (-0.59, 0.00)
Dimen. Change Card Sort	<b>0.33</b> <b>(0.15, 0.51)</b>	<b>0.34</b> <b>(0.10, 0.58)</b>	0.02 (-0.26, 0.29)	<b>0.34</b> <b>(0.13, 0.54)</b>	-0.01 (-0.25, 0.23)	-0.02 (-0.31, 0.27)
Flanker	0.05 (-0.13, 0.23)	<b>0.52</b> <b>(0.28, 0.75)</b>	<b>0.55</b> <b>(0.29, 0.82)</b>	<b>0.26</b> <b>(0.06, 0.47)</b>	0.25 (0.01, 0.49)	-0.27 (-0.55, 0.02)
List Sorting	0.07 (-0.11, 0.25)	0.21 (-0.02, 0.45)	0.12 (-0.15, 0.39)	<b>-0.24</b> <b>(-0.45, -0.04)</b>	-0.29 (-0.53, -0.05)	<b>-0.48</b> <b>(-0.77, -0.19)</b>
Pattern Comparison	0.18 (0.00, 0.36)	<b>0.63</b> <b>(0.39, 0.87)</b>	<b>0.45</b> <b>(0.18, 0.73)</b>	0.04 (-0.17, 0.25)	-0.14 (-0.39, 0.10)	<b>-0.57</b> <b>(-0.87, -0.28)</b>
Picture Sequence Memory	0.00 (-0.18, 0.19)	<b>0.38</b> <b>(0.14, 0.62)</b>	<b>0.34</b> <b>(0.06, 0.62)</b>	0.01 (-0.20, 0.23)	0.04 (-0.21, 0.29)	-0.31 (-0.61, -0.01)
Picture Vocabulary	0.03 (-0.14, 0.21)	-0.02 (-0.26, 0.21)	-0.07 (-0.34, 0.19)	<b>-0.40</b> <b>(-0.61, -0.20)</b>	<b>-0.39</b> <b>(-0.63, -0.15)</b>	<b>-0.38</b> <b>(-0.66, -0.09)</b>

<sup>a</sup>Effect sizes bolded where significant after multiple comparison correction using a 5% false discovery rate per Benjamini-Hochberg within each of the three outcome blocks separately for TBI severity (m=27/15/21) and TBI versus OTC comparisons (m=27/15/21).

Comparisons are propensity-weighted and adjusted for age, sex, and education.

30-SACT, 30 Seconds and Counting test of backward counting; BTACT, Brief Test of Adult Cognition by Telephone; GCS, admission Glasgow Coma Scale score; NIH Toolbox, National Institutes of Health Toolbox Cognition Battery; TMT, Trail Making Test; WAIS-IV PSI, Wechsler Adult Intelligence Scale-Fourth Edition Processing Speed Index.

processes, the BTACT RAVLT scores were less associated with type of injury (TBI, OTC) and TBI severity than the in-person RAVLT. Further, estimating the latent Episodic Memory dimension for the BTACT did little to enhance its convergent validity or sensitivity to TBI. This is not surprising given that only two measures contribute to the Episodic Memory factor, and one (im-

mediate recall) loads highly (0.92) onto the factor. Taken together, these findings imply that although the BTACT memory tests do appear to at least partly measure verbal episodic memory, relationships between the BTACT RAVLT indices and other variables (including but not limited to other cognitive performance metrics and TBI) are likely to be weaker than those measurable through a

more comprehensive in-person RAVLT assessment. The differential sensitivity of the BTACT versus in-person RAVLT to such factors must be taken into account, for example, when calculating power for TBI studies that utilize phone cognitive assessment or in deciding between phone versus traditional in-person neuropsychological testing. It may be valuable in future studies to explore whether adding more learning trials to the BTACT memory test would boost its performance.

In contrast to the BTACT Memory factor, which correlated more strongly with scores from the RAVLT than other tests, the EF factor and its subtests were robustly and comparably associated with performance in diverse cognitive domains (e.g., processing speed, set-shifting, receptive vocabulary). This may indicate that the EF factor reflects a more diffuse cognitive construct than is conveyed by its label.<sup>41</sup> Defined as processes that control and regulate thought and action, *executive functions* are typically considered to comprise moderate-complexity skills such as inhibiting prepotent responses, manipulating information in working memory, and set-shifting, or higher-level functions such as planning and reasoning.<sup>42,43</sup> Such executive functions are robustly correlated with other types of cognitive functions (e.g., general intellectual functioning, memory, or lower-level constructs such as processing speed)<sup>44</sup> but are generally viewed as distinct from these other constructs.<sup>42,45</sup>

In its aggregation of tasks intended to assess diverse executive functions (e.g., working memory manipulations, inductive reasoning) and other skills (e.g., processing speed), the BTACT's EF factor subtests tap, by design, a broad range of skills beyond pure executive functions. Further, classic neuropsychological tests such as those on which some BTACT subtests were based tend to be "impure" (i.e., rely on multiple cognitive skills).<sup>43</sup> Thus, it is not surprising that the BTACT EF (and General) factor behaves as a broad index with diffuse cognitive correlates. In the context of a broad cognitive screening tool, coverage of multiple cognitive domains is a feature that may increase the BTACT's chances of revealing cognitive dysfunction in the diverse TBI population. However, caution is warranted in interpreting the BTACT factors of Episodic Memory and Executive Functioning as reflecting only those functions, as those factors (and the constructs they represent) are multi-dimensional in nature.

It was reassuring that the 2-factor model was supported even though we did not include the optional Stop/Go response inhibition test used in prior publications that supported this model. This is promising, as scoring response latencies on the Stop/Go task requires recording audio files, which is not feasible in many settings. Further, we found that both the 2-factor and 1-factor model of the BTACT manifests strict measurement invariance across injury groups, including three TBI groups stratified by historical markers of injury severity and an OTC group. This indicates that differences between these groups in the BTACT factors can be interpreted as reflecting mean differences in the cognitive constructs reflected in each factor rather than other things that may contribute to differential cognitive performance across groups.

This is not to say, however, that the BTACT's factor structure is invariant across all groups or situations. Recent work by the instrument's authors suggests that the 2-factor structure is not invariant across a 9-year time interval, indicating that it may be problematic to interpret longitudinal changes in BTACT factor means as reflecting changes in the cognitive domains measured by the instrument.<sup>9</sup> The bifactor structure and other measurement approaches based in modern psychometrics that impose interval scaling properties may be better suited for the measurement of longitudinal change with the BTACT.<sup>13</sup> Future work should investigate whether the BTACT manifests longitudinal invariance

across differing time intervals within the trauma population specifically and consider its invariance across other groups of interest.

Regarding the relative sensitivity of different cognitive tests to TBI, there were cognitive indices from all three modes of assessment (phone, paper-and-pencil, and computerized) that demonstrated significantly poorer performance for the moderate-severe TBI versus OTC group, with maximum effect sizes in comparable ranges across modes of assessment ( $d \sim 0.5-0.6$ ). On average, however, effect sizes for moderate-severe TBI versus OTC group differences were larger for paper-and-pencil tests than the BTACT ( $Md -0.62$  vs.  $-0.43$ ).

Small but significant differences between mTBI and OTC groups were apparent for a limited number of indices across modes of assessment (e.g., BTACT General factor, EF factor, and TMT Part A,  $d \sim 0.2-0.3$ ), although effects for some NIH Toolbox tests were in the opposite direction from expected. Within the BTACT, the index most sensitive to TBI and TBI severity was the EF factor (which, as stated earlier, may be considered a general cognition factor). Differences between the BTACT and other tests were apparent when comparing different TBI severity strata, for which in-person tests yielded larger effect sizes than the BTACT when comparing mild (GCS 13-15) and moderate-severe (GCS 3-12) groups. The implication is that more comprehensive (and/or in-person) cognitive testing is likely preferable to abbreviated phone-based cognitive assessment, when feasible, for comparing participants of varying TBI severity. On the other hand, so long as its psychometric properties are well understood by examiners, the BTACT may be a valuable tool to facilitate cognitive assessment, particularly in situations in which it is often not attempted (i.e., phone calls).

In summary, we leveraged a large sample of trauma participants and orthopedic controls to reveal both strengths of the BTACT and the instrument's limitations, which likely stem, at least in part, from its abbreviated nature as compared with more traditional (in-person), comprehensive neuropsychological assessment. However, mode of assessment (phone vs. in-person) and length of assessment are conflated in this data set, making it unclear to what degree these factors contributed individually to the different performance of the BTACT and in-person tests. Additionally, it is unclear if the findings generalize beyond patients with TBI enrolled at Level 1 trauma centers who had acute head CT scans and were tested at 6 months post-injury. However, this investigation contributes substantial new knowledge about the construct validity of the BTACT and its utility in detecting TBI sequelae. Although more comprehensive neuropsychological assessment would likely yield stronger psychometric performance, the BTACT demonstrated a number of positive qualities that make it a promising tool for assessing patients with TBI, particularly when in-person or more comprehensive testing is infeasible.

### Acknowledgments

We thank Amy J. Markowitz, JD (Program Manager, TRACK-TBI/TBI Endpoints Development (TED) Traumatic Brain Injury Research Network, University of California, San Francisco) for providing editorial services for the article.

### The TRACK-TBI Investigators

Opeolu Adeoye, University of Cincinnati; Neeraj Badjatia, University of Maryland; Kim Boase, University of Washington; M. Ross Bullock, University of Miami; Randall Chesnut, University of Washington; John D. Corrigan, Ohio State University;

Karen Crawford, University of Southern California; Ramon Diaz-Arrastia, University of Pennsylvania; Ann-Christine Duhaime, Massachusetts General Hospital; Richard Ellenbogen, University of Washington; V. Ramana Feaser, Virginia Commonwealth University; Adam R. Ferguson, University of California, San Francisco; Brandon Foreman, University of Cincinnati; Raquel Gardner, University of California, San Francisco; Etienne Gaudette, University of Southern California; Dana Goldman, University of Southern California; Luis Gonzalez, TIRR Memorial Hermann; Shankar Gopinath, Baylor College of Medicine; Rao Gullapalli, University of Maryland; J. Claude Hemphill, University of California, San Francisco; Gillian Hotz, University of Miami; Sonia Jain, University of California, San Diego; C. Dirk Keene, University of Washington; Frederick K. Korley, University of Michigan; Joel Kramer, University of California, San Francisco; Natalie Kreitzer, University of Cincinnati; Chris Lindsell, Vanderbilt University; Joan Machamer, University of Washington; Christopher Madden, UT Southwestern; Amy Markowitz, University of California, San Francisco; Alastair Martin, University of California, San Francisco; Thomas McAllister, Indiana University; Randall Merchant, Virginia Commonwealth University; Laura B. Ngwenya, University of Cincinnati; David Okonkwo, University of Pittsburgh; Eva Palacios, University of California, San Francisco; Daniel Perl, Uniformed Services University; Ava Puccio, University of Pittsburgh; Miri Rabinowitz, University of Pittsburgh; Claudia Robertson, Baylor College of Medicine; Jonathan Rosand, Massachusetts General Hospital; Angelle Sander, Baylor College of Medicine; Gabriella Satris, University of California, San Francisco; David Schnyer, UT Austin; Seth Seabury, University of Southern California; Murray Stein, University of California, San Diego; Sabrina Taylor, University of California, San Francisco; Arthur Toga, University of Southern California; Alex Valadka, Virginia Commonwealth University; Mary Vassar, University of California, San Francisco; Paul Vespa, University of California, Los Angeles; Kevin Wang, University of Florida; John K. Yue, University of California, San Francisco; and Ross Zafonte, Harvard Medical School.

### Funding Information

The study was funded by the U.S. National Institutes of Health (NIH), National Institute of Neurological Disorders and Stroke (NINDS; Grants #U01 NS1365885 and #R01 NS110856), OneMind, and NeuroTrauma Sciences LLC. This work was also supported by the U.S. Department of Defense TED Initiative (Grant #W81XWH-14-2-0176). The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH and are not necessarily endorsed by the Department of Defense.

### Author Disclosure Statement

No competing financial interests exist.

### Supplementary Material

Supplementary Figure S1

### References

- Rabinowitz A.R., and Levin H.S. (2014). Cognitive sequelae of traumatic brain injury. *Psychiatr. Clin. North Am.* 37, 1–11.
- Whitnall L., McMillan T.M., Murray G.D., and Teasdale G.M. (2006). Disability in young people and adults after head injury: 5-7 year follow up of a prospective cohort study. *J. Neurol. Neurosurg. Psychiatry* 77, 640–645.
- Nelson L.D., Ranson J., Ferguson A.R., Giacino J., Okonkwo D.O., Valadka A.B., Manley G.T., McCrea M.A., and the TRACK-TBI Investigators. (2017). Validating multidimensional outcome assessment using the TBI Common Data Elements: an analysis of the TRACK-TBI Pilot sample. *J. Neurotrauma* 34, 3158–3172.
- Bashshur R.L., Shannon G.W., Krupinski E.A., Grigsby J., Kvedar J.C., Weinstein R.S., Sanders J.H., Rheuban K.S., Nesbitt T.S., Alverson D.C., Merrell R.C., Linkous J.D., Ferguson A.S., Waters R.J., Stachura M.E., Ellis D.G., Antoniotti N.M., Johnston B., Doarn C.R., Yellowlees P., Normandin S., and Tracy J. (2009). National telemedicine initiatives: essential to healthcare reform. *Telemed J. E Health* 15, 600–610.
- Johnston B., Wheeler L., Deuser J., and Sousa K.H. (2000). Outcomes of the Kaiser Permanente Tele-Home Health Research Project. *Arch. Fam. Med.* 9, 40–45.
- Marra D.E., Hamlet K.M., Bauer R.M., and Bowers D. (2020). Validity of teleneuropsychology for older adults in response to COVID-19: A systematic and critical review. *The Clinical neuropsychologist*, 1–42.
- Tun P.A., and Lachman M.E. (2006). Telephone assessment of cognitive function in adulthood: the Brief Test of Adult Cognition by Telephone. *Age Ageing* 35, 629–632.
- Lachman M.E., Agrigoroaei S., Tun P.A., and Weaver S.L. (2014). Monitoring cognitive functioning: psychometric properties of the brief test of adult cognition by telephone. *Assessment* 21, 404–417.
- Hughes M.L., Agrigoroaei S., Jeon M., Bruzzese M., and Lachman M.E. (2018). Change in cognitive performance from midlife into old age: findings from the Midlife in the United States (MIDUS) study. *J. Int. Neuropsychol. Soc.* 24, 805–820.
- U.S. Food and Drug Administration Pilot Clinical Outcome Assessment (COA) Compendium, Version 1. (2016). <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DevelopmentResources/UCM481225.pdf> (Last accessed October 18, 2019).
- Walton M.K., Powers J.H., Hobart J., Patrick D.L., Marquis P., Vamvakas S., and Isaac M. (2015). Clinical outcome assessments: conceptual foundation—report of the ISPOR Clinical Outcomes Assessment – Emerging Good Practices for Outcomes Research Task Force. *Value Health* 18, 741–752.
- Dams-O'Connor K., Sy K.T.L., Landau A., Bodien Y., Dikmen S., Felix E.R., Giacino J.T., Gibbons L., Hammond F.M., Hart T., Johnson-Greene D., Lengenfelder J., Lequerica A., Newman J., Novack T., O'Neil-Pirozzi T.M., and Whiteneck G. (2018). The feasibility of telephone-administered cognitive testing in individuals 1 and 2 years after inpatient rehabilitation for traumatic brain injury. *J. Neurotrauma* 35, 1138–1145.
- Gavett B.E., Crane P.K., and Dams-O'Connor K. (2013). Bi-factor analyses of the Brief Test of Adult Cognition by Telephone. *NeuroRehabilitation* 32, 253–265.
- Nelson L.D., Temkin N.R., Dikmen S., Barber J., Giacino J.T., Yuh E., Levin H.S., McCrea M.A., Stein M.B., Mukherjee P., Okonkwo D.O., Diaz-Arrastia R., Manley G.T., and the TRACK-TBI Investigators. (2019). Recovery after mild traumatic brain injury in patients presenting to US level I trauma centers: a Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) study. *JAMA Neurol.* 76, 1049–1059.
- Dikmen S., Machamer J., and Temkin N. (2017). Mild traumatic brain injury: longitudinal study of cognition, functional status, and post-traumatic symptoms. *J. Neurotrauma* 34, 1524–1530.
- Rey A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie* 28, 215–285.
- Wechsler D. (2008). *Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV)*. Pearson: Bloomington, MN.
- Reitan R.M. (1971). Trail making test results for normal and brain-damaged children. *Percept. Mot. Skills* 33, 575–581.
- Majdan A., Sziklas V., and Jones-Gotman M. (1996). Performance of healthy subjects and patients with resection from the anterior temporal lobe on matched tests of verbal and visuo-perceptual learning. *J. Clin. Exp. Neuropsychol.* 18, 416–430.
- Geffen G.M., Butterworth P., and Geffen L.B. (1994). Test-retest reliability of a new form of the auditory verbal learning test (AVLT). *Arch. Clin. Neuropsychol.* 9, 303–316.
- Weintraub S., Dikmen S.S., Heaton R.K., Tulsky D.S., Zelazo P.D., Bauer P.J., Carlozzi N.E., Slotkin J., Blitz D., Wallner-Allen K., Fox N.A., Beaumont J.L., Mungas D., Nowinski C.J., Richler J.,

- Deocampo J.A., Anderson J.E., Manly J.J., Borosh B., Havlik R., Conway K., Edwards E., Freund L., King J.W., Moy C., Witt E., and Gershon R.C. (2013). Cognition assessment using the NIH Toolbox. *Neurology* 80, S54–S64.
22. King N.S., Crawford S., Wenden F.J., Moss N.E., and Wade D.T. (1995). The Rivermead Post Concussion Symptoms Questionnaire: a measure of symptoms commonly experienced after head injury and its reliability. *J. Neurol.* 242, 587–592.
23. Derogatis L. R. (2001). *Brief Symptom Inventory 18 (BSI-18): Administration, Scoring, and Procedures Manual*. Pearson: Bloomington, MN.
24. Spitzer R.L., Kroenke K., and Williams J.B. (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 282, 1737–1744.
25. Kroenke K., Spitzer R.L., Williams J.B., and Lowe B. (2010). The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen. Hosp. Psychiatry* 32, 345–359.
26. Blevins C.A., Weathers F.W., Davis M.T., Witte T.K., and Domino J.L. (2015). The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. *J. Trauma. Stress* 28, 489–498.
27. von Steinbuechel N., Wilson L., Gibbons H., Muehlan H., Schmidt H., Schmidt S., Sasse N., Koskinen S., Sarajuuri J., Hofer S., Bullinger M., Maas A., Neugebauer E., Powell J., von Wild K., Zitnay G., Bakx W., Christensen A. L., Formisano R., Hawthorne G., and Truelle J.L. (2012). QOLIBRI Overall Scale: a brief index of health-related quality of life after traumatic brain injury. *J. Neurol. Neurosurg. Psychiatry* 83, 1041–1047.
28. Ware J., Jr., Kosinski M., and Keller S.D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med. Care* 34, 220–233.
29. Diener E., Emmons R.A., Larsen R.J., and Griffin S. (1985). The Satisfaction With Life Scale. *J. Pers. Assess.* 49, 71–75.
30. Jennett B., Snoek J., Bond M.R., and Brooks N. (1981). Disability after severe head injury: observations on the use of the Glasgow Outcome Scale. *J. Neurol. Neurosurg. Psychiatry* 44, 285–293.
31. Wilson J.T., Pettigrew L.E., and Teasdale G.M. (1998). Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. *J. Neurotrauma* 15, 573–585.
32. Hughes M.L., Agrigoroaei S., Jeon M., Bruzzese M., and Lachman M.E. (2018). Change in cognitive performance from midlife into old age: findings from the Midlife in the United States (MIDUS) study. *J. Int. Neuropsychol. Soc.* 24, 805–820.
33. Hu L., and Bentler P.M. (1999). Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Struc. Equ. Modeling* 6, 1–55.
34. Cheung G.W., and Resnold R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struc. Equ. Modeling* 9, 233–255.
35. Chen F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struc. Equ. Modeling* 14, 464–504.
36. Muthén L.K., and Muthén B.O. (1998-2017). *Mplus User's Guide*, 8th ed. Muthén & Muthén: Los Angeles, CA.
37. McCaffrey D.F., Ridgeway G., and Morral A.R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychol. Methods* 9, 403–425.
38. Benjamini Y., and Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc.: Series B (Statistical Methodology)* 57, 289–300.
39. Ivnik R.J., Malec J.F., Tangalos E.G., Petersen R.C., Kokmen E., and Kurland L.T. (1990). The Auditory-Verbal Learning Test (AVLT): norms for ages 55 and older. *Psychol. Assess.* 2, 304–312.
40. Ivnik R.J., Malec J.F., Smith G.E., Tangalos E.G., Petersen R.C., Kokmen E., and Kurland L.T. (1992). Mayo's older Americans normative studies: updated AVLT norms for ages 56 to 97. *Clin. Neuropsychol.* 6, 83–104.
41. Latzman R.D., and Markon K.E. (2010). The factor structure and age-related factorial invariance of the Delis-Kaplan Executive Function System (D-KEFS). *Assessment* 17, 172–184.
42. Friedman N.P., and Miyake A. (2017). Unity and diversity of executive functions: individual differences as a window on cognitive structure. *Cortex* 86, 186–204.
43. Miyake A., Friedman N.P., Emerson M.J., Witzki A.H., Howerter A., and Wager T.D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100.
44. Duncan J., Emslie H., and Williams P. (1996). Intelligence and the frontal lobe: the organization of goal-directed behavior. *Cognit. Psychol.* 30, 257–303.
45. Friedman N.P., Miyake A., Corley R.P., Young S.E., DeFries J.C., and Hewitt J.K. (2006). Not all executive functions are related to intelligence. *Psychol. Sci.* 17, 172–179.

Address correspondence to:

Lindsay D. Nelson, PhD

Medical College of Wisconsin

8701 West Watertown Plank Road

Milwaukee, WI 53226

USA

E-mail: linelson@mcw.edu