## Research Article

# Validity of the NIH toolbox cognitive battery in a healthy oldest-old 85+ sample

Sara A. Nolin[1] (ID), Hannah Cowart[1], Stacy Merritt[2], Katalina McInerney[2], P. K. Bharadwaj[4], Mary Kate Franchetti[4], David A. Raichlen[5], Cortney J. Jessup[4], G. Alex Hishaw[4], Emily J. Van Etten[4], Theodore P. Trouard[4], David S. Geldmacher[1], Virginia G. Wadley[1], Eric S. Porges[3], Adam J. Woods[3], Ron A. Cohen[3], Bonnie E. Levin[2], Tatjana Rundek[2], Gene E. Alexander[4] and Kristina M. Visscher[1] (ID)

[1]University of Alabama at Birmingham, Heersink School of Medicine, Evelyn F. McKnight Brain Institute, Birmingham, AL, USA, [2]University of Miami, Miller School of Medicine, Evelyn F. McKnight Brain Institute, Miami, FL, USA, [3]University of Florida, Evelyn F. and William L. McKnight Brain Institute, Gainesville, FL, USA, [4]University of Arizona, Evelyn F. McKnight Brain Institute, Tucson, AZ, USA and [5]University of Southern California, Los Angeles, CA, USA

## Abstract

**Objective:** To evaluate the construct validity of the NIH Toolbox Cognitive Battery (NIH TB-CB) in the healthy oldest-old (85+ years old). **Method:** Our sample from the McKnight Brain Aging Registry consists of 179 individuals, 85 to 99 years of age, screened for memory, neurological, and psychiatric disorders. Using previous research methods on a sample of 85 + y/o adults, we conducted confirmatory factor analyses on models of NIH TB-CB and same domain standard neuropsychological measures. We hypothesized the five-factor model (Reading, Vocabulary, Memory, Working Memory, and Executive/Speed) would have the best fit, consistent with younger populations. We assessed confirmatory and discriminant validity. We also evaluated demographic and computer use predictors of NIH TB-CB composite scores. **Results:** Findings suggest the six-factor model (Vocabulary, Reading, Memory, Working Memory, Executive, and Speed) had a better fit than alternative models. NIH TB-CB tests had good convergent and discriminant validity, though tests in the executive functioning domain had high inter-correlations with other cognitive domains. Computer use was strongly associated with higher NIH TB-CB overall and fluid cognition composite scores. **Conclusion:** The NIH TB-CB is a valid assessment for the oldest-old samples, with relatively weak validity in the domain of executive functioning. Computer use's impact on composite scores could be due to the executive demands of learning to use a tablet. Strong relationships of executive function with other cognitive domains could be due to cognitive dedifferentiation. Overall, the NIH TB-CB could be useful for testing cognition in the oldest-old and the impact of aging on cognition in older populations.

**Keywords:** aged 85 and over; neuropsychological tests; cognition; confirmatory factor analysis; construct validity; test development

(Received 6 December 2021; final revision 15 May 2022; accepted 19 May 2022; First Published online 14 October 2022)

## Introduction

The population of individuals within the oldest-old age range (85 years and older) is rapidly growing (Vincent & Velkoff, 2010). However, the lack of available data with a comprehensive assessment of cognitive functions in healthy agers over age 85 limits research in this age cohort. The developers of the NIH Toolbox Cognitive Battery (NIH TB-CB) limited their collection of normative data to those ages 3–85. Technology use in the cognitive testing environment is emerging, so ensuring the validity of using these new neuropsychological testing methods in the aging population is essential – especially since this oldest-old population may be less likely than other age groups to be comfortable with technology usage. Results of this study inform the use of the NIH TB-CB in future research in the oldest-old population.

The NIH TB-CB strives towards brevity, portability, and homogeneity in neurobehavioral assessment research through short tasks performed on an iPad (Gershon et al., 2013). The Cognitive Battery covers a wide range of cognitive domains, including executive functioning, episodic memory, language, processing speed, attention, and working memory (Gershon et al., 2013). The NIH Toolbox has been shown to be valid in diverse samples of varying age, language, race, ethnicity, gender, education, developmental disability, and neurological conditions (Carlozzi, Goodnight, et al., 2017; Flores et al., 2017; Hackett et al., 2018; Heaton et al., 2014; Hessl et al., 2016; Ma et al., 2021; Mungas et al., 2013; Tulsky et al., 2017; Weintraub, Dikmen, et al., 2013; Weintraub et al., 2014). For example, Mungas et al. (2014) examined younger and older age groups but only up to age 85. While the NIH TB-CB has been used in older

adult samples (O'Shea et al., 2018), to our knowledge, no data findings have been reported on the effectiveness of the NIH TB-CB as a battery to measure cognitive functions in healthy individuals over age 85.
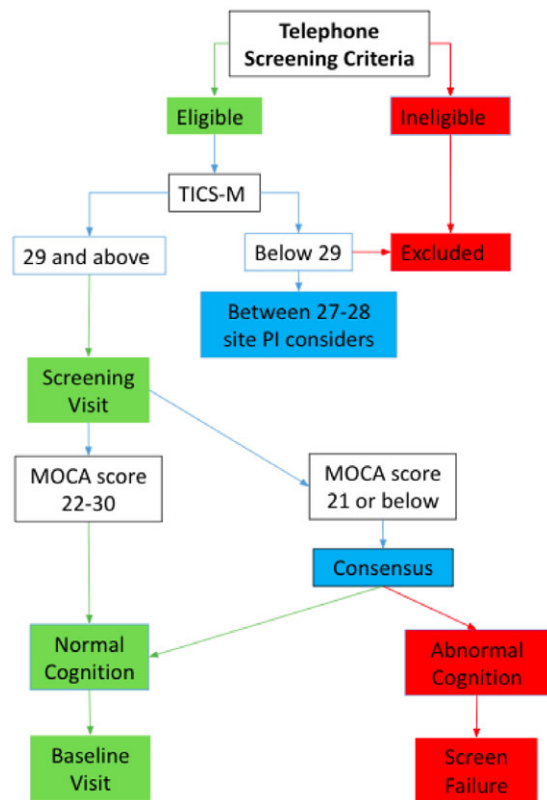
Factor analysis of the NIH Toolbox cognitive measures with standard neuropsychological tests of the same domains of cognition revealed good construct validity (Mungas et al., 2014), supporting correspondence between a given domain and a test used to measure it, which was indicated by the individual tests showing both strong associations with the hypothesized cognitive domains (i.e., convergent validity) and weak relationships between each of the tests and other domains (i.e., discriminant validity). In this case, Mungas et al. (2014) tested the validity of the NIH TB-CB in a cohort of adults, 20 to 85 years of age using confirmatory factor analysis. Although the NIH Toolbox assesses six specific domains (working memory, executive function, episodic memory, processing speed, vocabulary, and reading), they found that a five-factor model best describes the relationship between the NIH TB-CB and standard neuropsychological measures: Vocabulary, Reading, Episodic Memory, Working Memory, and Executive Function/Processing Speed with the NIH TB-CB tests falling largely within expected domains (Mungas et al., 2014). This factor structure did not vary across younger (ages 20–60) and older adults (ages 60–85), supporting the use of the NIH TB-CB as a measure of cognitive health across the adult age range from 20 to 85. Investigating factor structure of the NIH TB-CB in individuals older than age 85 provides an important opportunity to evaluate its utility for assessing cognitive functions in the fastest growing age group within the population of healthy older adults.

This study examines the validity of the NIH TB-CB cognitive domains in cognitively healthy older adults over age 85, which, to our knowledge, has yet to be reported. We employed a series of confirmatory factor analyses to investigate the convergent and discriminant validity, as well as the dimensional structure underlying the NIH TB-CHB and other validated measures of cognition in healthy older adults aged 85 years old and over. We hypothesized that the factor structure of the NIH Toolbox would be consistent across the lifespan, and thus the 5-factor model, derived from a younger adult sample (Mungas et al., 2014), would have a better model fit than alternative factor models. We also sought to evaluate the influence of demographic characteristics and computer use in this oldest-old age cohort on NIH TB-CB Composite scores.

## Method

### Participants

We analyzed data collected from the McKnight Brain Aging Registry, a cohort of community-dwelling, cognitively unimpaired older adults, 85 to 99 years of age. Figure 1 shows the extensive participant screening process. During initial screening over the phone, trained study coordinators administered the Telephone Interview for Cognitive Status modified (Cook et al., 2009) and an interview to assess for major exclusion criteria, which included individuals under age 85, severe psychiatric conditions, and neurological conditions, and cognitive impairment. Following the telephone screening, eligible participants underwent an in-person screening visit during which they were evaluated by a neurologist, a detailed medical history was obtained to assess health status and eligibility, and the Montreal Cognitive Assessment (MoCA) was administered (Nasreddine et al., 2005). An additional point was added for adjustment of the



**Figure 1.** Participant screening process. Telephone screening criteria included exclusion for major physical disabilities, dependence in instrumental activities of daily living or basic activities of daily living, uncontrolled medical conditions that would limit life expectancy or interfere with participation in the study, severe psychiatric conditions, neurological conditions (i.e., major vessel stroke, Parkinson's disease, dementia), active substance abuse or alcohol dependence, less than 6th-grade reading level, vision or hearing deficits that would cause impediment to cognitive test administration, MRI contraindications, and inability to follow study protocol and task instructions due to cognitive impairment. TICS-M was administered over the phone. An additional evaluation was included in the screening visit, including examination by a neurologist, geriatric depression scale, and detailed medical history.

MoCA score to account for non-white race and/or education equal to or below 12th grade. This adjustment was for the purpose of fairly screening individuals of lower education or non-white backgrounds and this adjustment is not based on normative data. The study was conducted in accordance with the Helsinki declaration. Approval for the study was received from the Institutional Review Boards at each of the data collection sites including University of Alabama at Birmingham, University of Florida, University of Miami, and University of Arizona.

Our fully screened sample consists of 192 community-dwelling individuals aged 85–99. We removed data from 13 participants from the analysis due to missingness related to administrator error, low visual acuity, participant's color blindness, or participant not completing the task. This left a remaining 179 participants in our sample. Only 138 participants were given the questionnaire related to computer use since this was adopted after data collection had begun; therefore analyses with the computer frequency variable are based on those 138 participants. Data from this group of healthy agers were collected using a standardized protocol across the four McKnight Institutes: University of Alabama at Birmingham, University of Florida, University of Miami, and University of Arizona. We recruited participants through mailings, flyers, physician referrals, and community-based recruitment.

**Table 1.** NIH toolbox measures and associated cognitive domains

| Measure | Scoring measure | Associated domains |
|---|---|---|
| Dimensional change card sort test (DCCS) | Computed scores | Executive function (cognitive flexibility and attention) |
| Flanker inhibitory control and attention | Computed scores | Executive function (cognitive flexibility and attention) |
| Picture sequence memory test | Theta scores | Episodic memory |
| Pattern comparison processing speed test | Raw scores | Attention and Processing Speed |
| List sorting working memory test | Raw scores | Working memory and executive functioning |
| Oral reading recognition test | Theta scores | Language (reading and crystalized abilities) |
| Picture vocabulary test | Theta scores | Language (receptive vocabulary) |

## Cognitive measures

Testing was performed by staff trained and certified across the four sites to administer the test battery. Testing was administered across two visits on separate days. We performed quality control on behavioral data through the double data entry tool in Redcap, wherein we entered data twice, and discrepancies were identified and corrected (Harris et al., 2019, 2009). The data were then again visually inspected and assessed for potential outliers and errors.

## NIH TB-CB measures

We used scores from the NIH TB-CB (Gershon et al., 2013; Weintraub et al., 2014), including the Dimensional Change Card Sort (DCCS) Test, the Flanker Inhibitory Control and Attention Test, the Picture Sequence Memory Test, the Pattern Comparison Processing Speed Test, the List Sorting Working Memory Test, the Oral Reading Recognition Test, and the Picture Vocabulary Test. The DCCSTest measures executive function by indicating a target characteristic and then instructing participants to quickly select the object that matches the indicated characteristic for that trial (either shape or color). The Flanker Inhibitory Control and Attention Test measures executive function by having participants quickly select the correct direction of an arrow among a set of arrows. The Pattern Comparison Processing Speed Test measures processing speed by having participants quickly decide whether or not two images match. The Picture Sequence Memory Test measures episodic memory by asking participants to place cards in a particular order from memory. The List Sorting Working Memory Test measures working memory by presenting an increasing number of pictures and then instructing participants to order the pictures by size and semantic category from memory. The Oral Reading Recognition Test measures language by having participants read aloud words shown on the screen. The Picture Vocabulary Test measures language by presenting a word verbally and instructing participants to select one of four images that best describes the word. Table 1 includes NIH TB-CB measures and their associated domains. Only raw or calculated scores were used for the analysis. We also performed follow-up analysis with demographically corrected scores for the NIH TB-CB scores.

## Standard neuropsychological measures

We used standard neuropsychological tests with strong psychometric properties within the same domains as those used in the NIH TB-CB. Memory functioning was assessed through the California Verbal Learning Test II (Delis et al., 1987), a word list learning task, and the Benson Figure Test (Beekly et al., 2007), a visual memory task. Executive functioning was assessed through the Trail Making Test B (Gaudino et al., 1995), visual attention and switching task; Wechsler adult intelligence scale-fourth edition

(WAIS-IV) Matrix Reasoning (Benson et al., 2010) subtest, which involves recognizing and utilizing pattern recognition and integration; and the Stroop Color Word-Inhibition test (MacLeod, 1991; MacLeod, 1992), an inhibition task. Language/Vocabulary was assessed through the WAIS-IV Similarities (Benson et al., 2010), which involves explaining abstract relationships between two words. Processing speed was assessed through the WAIS-IV Coding and Symbol Search (Benson et al., 2010) subtests which both involve speeded visual processing. Lastly, working memory was assessed through the WAIS-IV Letter-Number sequencing subtest (Benson et al., 2010), a task involving sequencing a set of letters and numbers, and Digit Span (Beekly et al., 2007), a number recall task including recall backward. Table 2 includes the standard neuropsychological measures and their associated domains. Only raw or calculated scores were used. We also performed follow-up analysis with demographically corrected scores for the NIH TB-CB scores.

## Confirmatory factor analysis

Based on the methods of previous work from Mungas et al. (2014), we performed a series of confirmatory factor analyses, which allowed us to assess the degree to which the original conceptual model of the NIH TB-CB aligns with the factor structure of the NIH TB-CB and standard neuropsychological measures of the same cognitive domains within the oldest-old. We compared models matching this conceptual model, as well as alternative models, detailed in Table 3.

Following Mungas et al. (2014), we included the following tests of model fit: overall Chi-square test of model fit as well as the Tucker Lewis Index (Tucker & Lewis, 1973), Comparative Fit Index (Bentler & Bonett, 1980; Bentler, 1990), the root mean square error of approximation (Browne & Cudeck, 1992), and Standardized Root Mean Square Residual (Bentler, 1989). We evaluated modification indices to see if there could be any significant improvement in the model by changing model parameters. We compared models using the Akaike Information Criterion (AIC). This approach can further establish the reproducibility of previous findings (Mungas et al., 2014) while extending it to our oldest-old cohort. We used R and the lavaan package to perform confirmatory factor analysis (Rosseel, 2012).

## Evaluation of validity

Convergent validity was assessed by examining factor loadings of NIH TB-CB on their domain factor and evaluating the correlation between an average of the standard neuropsychological measures of a domain and the NIH TB-CB of the domain. Discriminant validity was assessed by examining modification indices cross-loadings of NIH TB-CB measures, identifying high inter-correlation of factors, and evaluating the correlation between an average of

**Table 2.** Standard neuropsychological measures and associated cognitive domains

| Measures | Scoring measure | Domains |
|---|---|---|
| Trail making test (TMT) part B | Time (number of seconds per line drawn) | Executive functioning (switching) |
| Letter-number sequencing | Number of correct trials | Working memory (mental manipulation) |
| Digit span backward | Number of correct trials | Working memory (mental manipulation) |
| California verbal learning test (CVLT) | Total number of correct responses at delayed recall | Episodic verbal learning and memory |
| Matrix reasoning (WAIS-IV) | Total number of correct responses | Perceptual reasoning (executive function) |
| Coding (WAIS-IV) | Total number correct within the specified time limit | Processing speed |
| Symbol search (WAIS-IV) | The difference in number of correct responses and number of incorrect responses | Processing speed |
| Similarities (WAIS-IV) | Total number of correct responses | Verbal reasoning and comprehension |
| Stroop color-word inhibition | Interference score | Executive functioning (inhibition) |
| Benson figure test | Total score based on accuracy and placement of figure components | Episodic visual memory |

the standard neuropsychological measures of a domain and the NIH TB-CB of a different domain. These methods of assessing validity of cognitive measures have been applied previously (Andresen, 2000; Carlozzi, Tulsky, et al., 2017; Heaton et al., 2014; Tulsky et al., 2017; Weintraub, Bauer, et al., 2013).

### Multiple regression with NIH TB-CB composite scores

Composite scores are automatically generated through the NIH TB-CB. Prior work that generated normative data indicated that there is a decline of fluid composite scores with age and a plateau of crystallized composite scores after middle age (Casaletto et al., 2015). The composite scores from our sample fit with this trend (Table 4) with relatively similar crystallized scores as other older adults and lower fluid scores than younger adults age groups (Casaletto et al., 2015). Three multiple linear regressions predicted the three NIH TB-CB Uncorrected Composite Standard Scores-Total, Crystallized, and Fluid. Predictors included years of education, age, gender (1 = Male, 2 = Female), race (1 = White, 2 = Black/African American, 3 = Asian), and computer use frequency (0 = No computer experience/Not used a computer in last 3 months, 1 = Less than 1 hr a week, 2 = 1 hr but less than 5 hr a week, 3 = 5 hr but less than 10 hr a week, 4 = 10 hr but less than 15 hr a week, 5 = At least 15 hr a week). Table 4 includes descriptive statistics for these variables.

### Results

#### Model fit

Based on prior studies, we hypothesized that the 5-factor model of the NIH TB-CB and standard neuropsychological measures would have a better fit than alternative factor models. We found that the 5-factor (Language, Memory, Working Memory, Executive, and Speed) and 6-factor (Vocabulary, Reading, Memory, Working Memory, Executive, and Speed) models have similar fit indices that indicate good fit (Table 5). The 6-factor model had a slightly smaller AIC (5-factor AIC = 16608.769 and 6-factor AIC = 16606.818). We, therefore, chose the 6-factor model as the best fit. The model aligns with the original six domains of the NIH TB-CB (working memory, executive function, episodic memory, processing speed, language, and reading) (Gershon et al., 2013).

The 5-factor model found in Mungas et al. (2014) (Vocabulary, Reading, Memory, Working Memory, Executive/Speed) was

different from the 5-factor model our study found to be a good fit in the oldest-old sample. Mungas et al. (2014) found that the model that combined executive and speed into one factor, and separated vocabulary and reading into two separate factors, was a better fit than a 5-factor model that separated executive and speed factors and instead combined vocabulary and reading into a language factor (Table 5). We did not find an inter-correlation between executive and speed factors>.9 as was found in prior studies (Mungas et al., 2014; Tulsky et al., 2017).

#### Convergent validity

Standardized coefficients for the 6-factor model (Table 6) showed NIH TB-CB measures loaded strongly on their respective factors, supporting convergent validity. Picture Vocabulary loaded very highly (.82) on the Vocabulary factor; List Sorting loaded highly (.634) on the Working Memory factor; both DCCS and Flanker loaded strongly (.62 and .585) on the Executive Functioning factor; Picture Sequencing loaded strongly (.533) on the Memory factor, and Pattern Comparison had a moderate loading (.442) on the Speed factor. When we analyzed the correlation between an average of the standard neuropsychological measures of a domain and the NIH TB-CB of the domain, we found Picture Sequence, List Sorting, Pattern Comparison, and Picture Vocabulary all had adequate correlations for convergent validity; however, Flanker and DCCS measures had weak correlations with the executive functioning standard neuropsychological measures, and therefore convergent validity was not supported based on this metric (Andresen, 2000). Since we did not have a standard neuropsychological measure available to load with the NIH TB-CB Oral Reading Task, convergent validity for the Reading domain was not assessed.

#### Discriminant validity

Only one weak modification index indicated a split loading of the NIH TB-CB Flanker measure on the Vocabulary factor. The lack of strong cross-loadings between factors indicated discriminant validity of our model. Additionally, the inter-correlations among the six factors (inter-correlation range of $r = .157–.811$; Table 7) were within acceptable limits as used in a prior NIH TB validity study (Tulsky et al., 2017). The consistently highest inter-correlations were between executive functioning and other domains (inter-correlations range of $r = .428–.811$; Table 7). While vocabulary and reading were correlated ($r = .641$), these two crystallized

**Table 3.** Alternative models

| | | 1 Factor model |
|---|---|---|
| Model | Factors | Measures used in each factor |
| 1a | Global | All |
| **2 Factor models** | | |
| Model | Factors | Measures used in Each Factor |
| 2a | Crystallized | Oral reading recognition, picture vocabulary, similarities |
| | Fluid | Picture sequence memory, list sorting WM, LNS, digit span backward, CVLT, Benson figure, DCCS, pattern comparison processing speed, list sorting WM, TMT part B, matrix reasoning, coding, symbol search, Stroop color-word inhibition, flanker |
| 2b | Memory | Picture sequence memory, list sorting WM, LNS, digit span backward, CVLT, Benson figure |
| | Non-memory | DCCS, pattern comparison processing speed, list sorting WM, TMT part B, matrix reasoning, coding, symbol search, Stroop color-word inhibition, flanker |
| **3 Factor models** | | |
| Model | Factors | Measures used in each factor |
| 3a | Language | Oral reading recognition, picture vocabulary |
| | EM/WM | Picture sequence memory, list sorting WM, LNS, digit span backward, CVLT, Benson figure |
| | EF/speed | DCCS, pattern comparison processing speed, list sorting WM, TMT part B, matrix reasoning, coding, symbol search, Stroop color-word inhibition, flanker |
| 3b | Language | Oral reading recognition, picture vocabulary |
| | EM | Picture sequence memory, CVLT, Benson figure |
| | WM/EF/Speed | DCCS, pattern comparison processing speed, list sorting WM, TMT part B, LNS, digit span backward, matrix reasoning, coding, symbol search, Stroop color-word inhibition, flanker |
| **4 Factor Models** | | |
| Model | Factors | Measures used in each factor |
| 4a | Language | Oral reading recognition, picture vocabulary |
| | EM | Picture sequence memory, Benson figure |
| | WM | List sorting WM, LNS, digit span backward |
| | EF/Speed | DCCS, pattern comparison processing speed, list sorting WM, TMT part B, matrix reasoning, coding, symbol search, Stroop color-word inhibition, flanker |
| 4b | Vocabulary | Picture Vocabulary |
| | Reading | Oral reading recognition |
| | EM | Picture sequence memory, CVLT, Benson figure |
| | WM/EF/Speed | DCCS, pattern comparison processing speed, list sorting WM, trail making part B, LNS, digit span backward, matrix reasoning, symbol search and coding, Stroop color-word inhibition, flanker |
| 4c | Vocabulary | Picture vocabulary |
| | Reading | Oral reading recognition |
| | EM/WM | Picture sequence memory, list sorting WM, LNS, digit span backward, CVLT, Benson figure |
| | EF/speed | DCCS, pattern comparison processing speed, list sorting WM, TMT part B, matrix reasoning, coding, symbol search, Stroop color-word inhibition, flanker |
| **5 Factor models** | | |
| Model | Factors | Measures used in each factor |
| 5a | Language | Oral reading recognition, picture vocabulary |
| | EM | Picture sequence memory, CVLT, Benson figure |
| | WM | List sorting WM, LNS, digit span backward |
| | EF | DCCS, list sorting WM, TMT Part B, matrix reasoning, Stroop color-word inhibition, flanker |
| | Speed | Pattern comparison processing speed, coding, symbol search |
| 5b | Vocabulary | Picture vocabulary |
| | Reading | Oral reading recognition |
| | EM | Picture sequence memory, CVLT, Benson figure |
| | WM | List sorting WM, LNS, digit span backward |
| | EF/speed | DCCS, pattern comparison processing speed, list sorting WM, TMT part B, matrix reasoning, coding, symbol search, Stroop color-word inhibition, flanker |
| **6 Factor model** | | |
| Model | Factors | Measures used in each factor |
| 6a | Vocabulary | Picture vocabulary |
| | Reading | Oral reading recognition |
| | EM | Picture sequence memory, CVLT, Benson figure |
| | WM | List sorting WM, digit span backward |
| | EF | DCCS, list sorting WM, TMT part B, matrix reasoning, Stroop color-word inhibition, flanker |
| | Speed | Pattern comparison processing speed, coding, symbol search |

intelligence factors were also related to fluid intelligence factors; therefore, there was no clear crystallized/fluid separation.

The correlation between an average of the standard neuropsychological measures of a domain and the NIH TB-CB of a different domain indicated Picture Sequence, List Sorting, Pattern Comparison, and Picture Vocabulary all had correlations to other domains that were smaller than the correlation to their domain (Figure 2). However, Flanker and DCCS measures had higher correlations to domains outside Executive Functioning; therefore, this metric indicates poor discriminant validity for these measures.

Together, convergent and discriminant validity evidence indicates sufficient construct validity of the NIH TB-CB within an 85+ cohort, with relatively weaker construct validity for executive functioning measures in the NIH TB-CB.

**Table 4.** Participant characteristics

| Participant characteristics total sample ($n = 179$) | |
|---|---|
| Age (Years), $+/-SD$ (Years) | 88.34 $+/-$3.06 (85–99) |
| Education (Years), Mean $+/-SD$ (range) | 18.13$+/-$2.69 (11–22) |
| Race and Education Adjusted MoCA, Mean $+/-SD$ (range) | 24.78 $+/-$2.51 (17–30) |
| NIH TB Cognitive Total Composite demographically corrected* standard score, Mean $+/-SD$ (range) | 104.75$+/-$12.18 (71–135) |
| NIH TB Cognitive Fluid Composite demographically corrected* standard score, Mean $+/-SD$ (range) | 97.52 $+/-$12.7 (75–130) |
| NIH TB Cognitive Crystallized Composite demographically corrected* standard score, Mean $+/-SD$ (range) | 110.7 $+/-$12.84 (71–170) |
| Sex, *n* (%) | |
| Female | 96 (53.63%) |
| Male | 83 (46.36%) |
| Race/ethnicity, *n* (%) | |
| Non-Hispanic Caucasian | 165 (92.21%) |
| African American | 6 (3.35%) |
| Hispanic Caucasian | 5 (2.79%) |
| Asian | 3 (1.67%) |
| Marital status, *n* (%) | |
| Widowed | 87 (48.6%) |
| Married | 70 (39.1%) |
| Divorced | 13 (7.26%) |
| Domestic partnership | 6 (3.35%) |
| Never married | 3 (1.67%) |
| Sample subset ($n = 138$) | |
| Computer use frequency, Mean $+/-SD$ (range) | 2.8$+/-$1.75(0–5) |
| NIH TB uncorrected total cognition composite score, Mean $+/-SD$ (range) | 95 $+/-$8.53 (76–116) |
| NIH TB uncorrected fluid cognition composite score, Mean $+/-SD$ (range) | 80$+/-$9.38 (61–102) |
| NIH TB uncorrected crystallized cognition composite score, Mean $+/-SD$ (range) | 112$+/-$7.87 (93–129) |

*Note that demographic corrections are not available for individuals over age 85; therefore, corrections for all participants, including those over 85 years of age, were based on normative data for individuals aged 85 years old.

**Table 5.** Model fit indices

| Model | Overall $\chi^2$ [df] | CFI | TLI | RMSEA (90% CI) | SRMR | AIC |
|---|---|---|---|---|---|---|
| 1a | 283.765 [119] | 0.784 | 0.753 | 0.088 (0.075–0.101) | 0.075 | 16699.98 |
| 2a | 240.598 [118] | 0.839 | 0.814 | 0.076 (0.062–0.09) | 0.072 | 16658.794 |
| 2b | 274.544 [118] | 0.794 | 0.763 | 0.086 (0.073–0.099) | 0.074 | 16692.739 |
| 3a | 223.483 [116] | 0.859 | 0.834 | 0.072 (0.058–0.086) | 0.071 | 16645.678 |
| 3b | 220.664 [116] | 0.862 | 0.839 | 0.071 (0.057–0.085) | 0.068 | 16642.86 |
| 4a | 195.412 [113] | 0.892 | 0.87 | 0.064 (0.048–0.079) | 0.065 | 16623.608 |
| 4b | 215.099 [114] | 0.867 | 0.842 | 0.070 (0.056–0.085) | 0.067 | 16641.294 |
| 4c | 221.732 [114] | 0.858 | 0.831 | 0.073 (0.058–0.087) | 0.07 | 16647.928 |
| 5a | 172.573 [109] | 0.916 | 0.896 | 0.057(0.040–0.073) | 0.06 | 16608.769 |
| 5b | 187.030 [110] | 0.899 | 0.875 | 0.063 (0.047–0.078) | 0.064 | 16621.225 |
| 6a | 162.623 [105] | 0.924 | 0.902 | 0.055 (0.038–0.072) | 0.058 | 16606.818 |

Overall chi-squared measures how well a model compares to observed data. Comparative Fit Index (CFI) examines the discrepancy between data and the hypothesized model. Tucker Lewis Index (TLI) analyzes the discrepancy between the x² of the hypothesized model and the null model. The Root Mean Squared Error of Approximation (RMSEA) analyzes discrepancy between the hypothesized model (with optimal parameter estimates) and population covariance matrix. The Standardized Root Mean Square Residual (SRMR) is the root of the discrepancy between the sample covariance matrix and model covariance matrix. The Akaike Information Criterion (AIC) is a value used to evaluate how well a model fits the data. Lower is a better fit.

### *Predictors of NIH TB-CB composite scores*

Race ($\beta = -3.5$, $p = .009$), data collection site ($\beta = 1.98$, $p = .017$), computer use frequency ($\beta = 1.19$, $p = .007$), and years of education ($\beta = .64, p = .021$) were significant predictors of the NIH TB-CB Total composite score and the overall model's adjusted R² was .1541 ($p < .001$) There was a significant $R^2$-change of .164 ($p < .001$) between the first block of the covariate, site, and the second block with race, gender, age, years of education, and computer use frequency.

Only computer use frequency ($\beta = 1.12$, $p = .02$) was a significant predictor of NIH TB-CB Fluid composite score, and the overall model's adjusted R² was .03 ($p = .07$). There was a significant $R^2$-change of .0717 ($p = .042$) between the first block of the covariate, site, and the second block with race, gender, age, years of education, and computer use frequency.

Race ($\beta = -4.16$, $p = .001$) and years of education ($\beta = 1.1, p < .001$) were significant predictors of NIH TB-CB Crystallized composite score, and the overall model's adjusted $R^2$ was .21 ($p < .001$). There was a significant $R^2$-change of .17 ($p < .001$) between the first block of the covariate, site, and the second block with race, gender, age, years of education, and computer use frequency.

### *Follow-up analysis with demographically corrected scores for the NIH TB-CB*
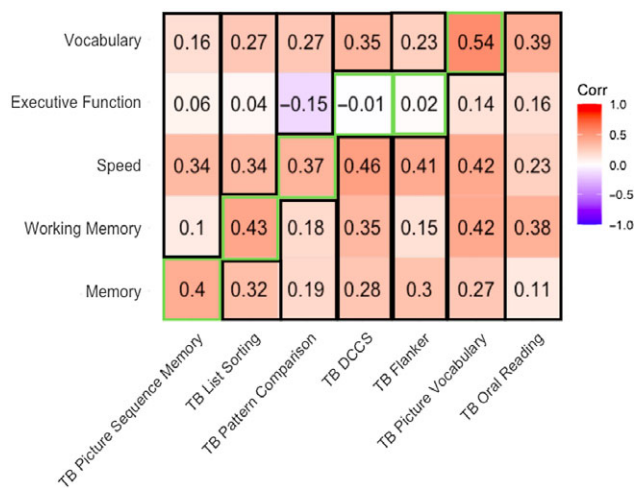
We repeated the analysis with demographically corrected scores for the NIH TB-CB and found no differences in the interpretation of the findings including no changes in determination of best-fitting model, validity or predictors of NIH TB-CB Composite Scores

**Table 6.** Standardized coefficients for the 6-factor model

| Latent factor | Observed indicator | Loading |
|---|---|---|
| Vocabulary | TB picture vocabulary | 0.82 |
| | WAIS-IV similarities | 0.664 |
| Reading | TB oral reading | 1 |
| Episodic memory | TB picture sequence memory | 0.533 |
| | CVLT | 0.703 |
| | Benson figure | 0.413 |
| Working memory | TB list sorting | 0.634 |
| | Letter-number sequence | 0.503 |
| | Digit span backwards | 0.648 |
| Executive function | TB DCCS | −0.62 |
| | TB Flanker test | −0.585 |
| | WAIS-IV matrix reasoning | −0.555 |
| | Stroop interference | −0.293 |
| | Trails B | 0.675 |
| Speed | TB pattern comparison | 0.442 |
| | WAIS-IV coding | 0.72 |
| | WAIS-IV symbol search | 0.807 |

**Table 7.** Inter-correlation of factors for the 6-factor model

| | Episodic memory | Working memory | Speed | Executive function | Vocabulary |
|---|---|---|---|---|---|
| Working memory | 0.487 | | | | |
| Speed | 0.587 | 0.499 | | | |
| Executive function | −0.616 | −0.77 | −0.811 | | |
| Vocabulary | 0.537 | 0.68 | 0.579 | −0.715 | |
| Reading | 0.157 | 0.494 | 0.238 | −0.428 | 0.641 |



**Figure 2.** Correlations between NIH TB-CB measures and standard neuropsychological domain average. TB = NIH TB-CB measures; green outline = within domain correlation, consistent with convergent validity; black outline = outside domain correlation, consistent with discriminant validity.

(see Supplemental Table 1 and 2 for model fit indices and model standardized coefficients).

## Discussion

In our cohort of cognitively unimpaired, older adults over 85 years of age, the NIH TB-CB tests and standard neuropsychological measures had convergent and discriminant validity, consistent with the six domains of cognition initially intended to be evaluated by the NIH TB-CB. These findings suggest the NIH TB-CB has construct validity in oldest-old adults, ages 85–99. The 5-factor model (model 5a), which combines reading and vocabulary into a language factor, also displayed a good model fit. There was relatively less evidence to support the combination of executive function and speed factors, as shown in the 5-factor model by Mungas et al. (2014). However, there were strong relationships between executive function and all other factors. We also found that computer use frequency strongly predicted the total and fluid NIH TB-CB composite scores, suggesting that either: (1) greater experience with computers impacts performance on this tablet-based assessment; or (2) having lower cognitive capacity (reflected in the NIH TB-CB scores) leads to less computer use.

### Cognitive dedifferentiation and the executive decline hypothesis

Cognitive dedifferentiation describes the tendency for separable cognitive abilities (such as language and executive function) to become less separable with age; dedifferentiation may reflect underlying cognitive impairment (Baltes et al., 1980; Batterham et al., 2011; Hülür et al., 2015; Wallert et al., 2021; Wilson et al., 2012). Our findings of more widespread domain inter-correlations with executive functioning could reflect age-related cognitive dedifferentiation. We only included healthy individuals in our sample, so this cognitive dedifferentiation may be a result of healthy aging.

A potential explanation for the strong relationship between executive function and other cognitive domains is that executive functions may play a greater role in supporting nonexecutive task performance in older people, as outlined in the executive decline hypothesis (Crawford et al., 2000; Ferrer-Caja et al., 2002; Salthouse et al., 2003). Prior factor analysis research has shown similar relationships between executive and nonexecutive tasks (Lamar et al., 2002). This reflects on a broader issue in classifying tests as measuring only a single domain. The NIH TB was intentionally developed so that each cognitive domain would be linked to one or two tasks from the toolbox. The domains are not pure, however, and the tests used to assess each are likely to be affected by performance limitations in other domains.

We found that the best-fitting model was the 6 factor model, rather than the 5 factor model (model 5b) that Mungas et al. (2014) found to best fit data for a younger sample. The difference in best-fit model could be due to increased associations of executive function with all other domains in the oldest-old. Both the hypothesis of greater cognitive dedifferentiation with age and the executive decline hypothesis would predict increased association of executive function with other domains, as was observed. Thus, our result is likely to represent a more holistic effect than simply reflecting a tight coupling between executive functioning and speed domains in this population.

### Role of computer use frequency in cognitive performance

Younger age, higher education, non-Hispanic ethnicity, physical health, and mental health have been shown to be predictors of greater computer use (Werner et al., 2011). Additionally, perceptual speed moderates the relationship between age and technology ownership (Kamin & Lang, 2016). Previous work has indicated a relationship between cognitive performance and the level of computer experience (Fazeli et al., 2013; Wu et al., 2019). Since computer use frequency was a significant predictor of total and fluid composite scores, technological familiarity may play a crucial role

in performance on NIH TB-CB measures. Participants who were relatively less familiar with technology may have also experienced increased demand on executive functioning as they had to learn technological skills while also performing a cognitive task. Alternatively, since adept use of computers requires cognitive abilities such as executive functioning and processing speed, participants with lower cognitive abilities may tend to avoid engagement with computers in their daily lives due to the cognitive demands of computer use.

Additionally, there are key differences between paper-and-pencil tasks and tablet-based tasks that could impact performance, such as less ability to self-correct, less flexibility for the administrator to pace the task appropriately for the participant, and less engagement between the administrator and the participant (Aşkar et al., 2012). Attitudes towards computers could have resulted in a lower frequency of computer use and, therefore, a negative impact on their cognitive scores (Fazeli et al., 2013). Future work should investigate participants' disposition towards computers and their current computer use. This could impact the usability of the NIH TB-CB in older samples since older adults are less likely to have familiarity with technology than younger cohorts (Victorson et al., 2013; Werner et al., 2011). Researchers may need to assess a participant's technology use to determine the appropriateness of using the NIH TB-CB. Alternatively, composite scores could account for current and past computer use in the calculation of standardized scores (Lee Meeuw Kjoe et al., 2021).

### Limitations

This study has limitations. We did not have a standard neuropsychological measure similar to the Oral Reading test available in the dataset, so convergent validity for that factor could not be fully tested in our study. Our sample is also mostly white and highly educated, which limits the generalizability of this work. We also acknowledge that in our confirmatory factor analysis, we could not account for variability that may have occurred across data collection sites. However, substantial efforts were made to homogenize data collection across sites and we included site as a covariate in our regression models (Section "Predictors of NIH TB-CB Composite Scores"). Future work could further describe the oldest-old cohort through comparisons of this sample to other age cohorts who also completed the iPad version of the NIH TB-CB.

### Conclusions

The NIH TB-CB was created to solve issues of inconsistency and difficulty of administration of neuropsychological testing in research, focusing on those ages 3 to 85. Our findings suggest that this test battery could be valuable for the assessment of cognitive health in individuals over 85. Having a common metric on an easy-to-use iPad tablet could enable research studies to include larger and more representative samples of older adults, and researchers could easily compare these scores to other studies which have adopted the NIH TB-CB. The NIH TB-CB could also be helpful since it can provide precise timing metrics along with cognitive accuracy scores for use in aging studies. This work has confirmed the construct validity and the feasibility of the NIH TB-CB in an 85+ sample, which will provide a basis for the usability of the battery in future older adult research. However, there may be limitations in the NIH TB-CB's ability to validly assess individuals with low computer use and validly measure executive functioning,

possibly due to age-related changes in executive functioning's relationship with other cognitive domains. This work provides a pathway towards broadening the age span of the NIH TB-CB to 99 years of age which will allow longitudinal and cohort studies to compare across almost the entire human lifespan (3–99).

### References

Andresen, E. M. (2000). Criteria for assessing the tools of disability outcomes research. *Archives of Physical Medicine and Rehabilitation*, *81*, S15–S20. https://doi.org/10.1053/apmr.2000.20619

Aşkar, P., Altun, A., Cangöz, B., Cevik, V., Kaya, G., & Türksoy, H. (2012). A comparison of paper-and-pencil and computerized forms of line orientation and enhanced cued recall tests. *Psychological Reports*, *110*, 383–396. https://doi.org/10.2466/03.22.PR0.110.2.383-396

Baltes, P. B., Cornelius, S. W., Spiro, A., Nesselroade, J., & Willis, S. (1980). Integration *versus* differentiation of fluid/crytallized intelligence in old age. *Developmental Psychology*, *16*, 625–635. https://doi.org/10.1037/0012-1649.16.6.625

Batterham, P. J., Christensen, H., & Mackinnon, A. J. (2011). Comparison of age and time-to-death in the dedifferentiation of late-life cognitive abilities. *Psychology and Aging*, *26*, 844–851. https://doi.org/10.1037/a0023300

Beekly, D. L., Ramos, E. M., Lee, W. W., Deitrich, W. D., Jacka, M. E., Wu, J., Hubbard, J. L., Koepsell, T. D., Morris, J. C., Kukull, W. A., & NIA Alzheimer's Disease Centers. (2007). The National Alzheimer's coordinating center (NACC) database: The uniform data set. *Alzheimer Disease and Associated Disorders*, *21*, 249–258. https://doi.org/10.1097/WAD.0b013e318142774e

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler adult intelligence scale-fourth edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment*, *22*, 121–130. https://doi.org/10.1037/a0017767

Bentler, P. M. (1989). *EQS 6 structural equations program manual*. Multivariate Software, Inc.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246. https://doi.org/10.1037/0033-2909.107.2.238

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

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*, 230–258. https://doi.org/10.1177/0049124192021002005

Carlozzi, N. E., Goodnight, S., Casaletto, K. B., Goldsmith, A., Heaton, R. K., Wong, A. W. K., Baum, C. M., Gershon, R., Heinemann, A. W., & Tulsky, D. S. (2017). Validation of the NIH toolbox in individuals with neurologic disorders. *Archives of Clinical Neuropsychology*, *32*, 555–573. https://doi.org/10.1093/arclin/acx020

Carlozzi, N. E., Tulsky, D. S., Wolf, T. J., Goodnight, S., Heaton, R. K., Casaletto, K. B., Wong, A. W. K., Baum, C. M., Gershon, R. C., & Heinemann, A. W. (2017). Construct validity of the NIH toolbox cognition battery in individuals with stroke. *Rehabilitation Psychology*, *62*, 443–454. https://doi.org/10.1037/rep0000195

Casaletto, K. B., Umlauf, A., Beaumont, J., Gershon, R., Slotkin, J., Akshoomoff, N., & Heaton, R. K. (2015). Demographically corrected normative standards for the English version of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society*, 21, 378–391. https://doi.org/10.1017/S1355617715000351

Cook, S. E., Marsiske, M., & McCoy, K. J. M. (2009). The use of the modified telephone interview for cognitive status (TICS-M) in the detection of amnestic mild cognitive impairment. *Journal of Geriatric Psychiatry and Neurology*, 22, 103–109. https://doi.org/10.1177/0891988708328214

Crawford, J. R., Bryan, J., Luszcz, M. A., Obonsawin, M. C., & Stewart, L. (2000). The Executive decline hypothesis of cognitive aging: Do executive deficits qualify as differential deficits and do they mediate age-related memory decline? *Aging, Neuropsychology, and Cognition*, 7, 9–31. https://doi.org/10.1076/anec.7.1.9.806

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California verbal learning test* (2nd ed.). Psychological Corporation.

Fazeli, P. L., Ross, L. A., Vance, D. E., & Ball, K. (2013). The relationship between computer experience and computerized cognitive test performance among older adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 68, 337–346. https://doi.org/10.1093/geronb/gbs071

Ferrer-Caja, E., Crawford, J. R., & Bryan, J. (2002). A structural modeling examination of the executive decline hypothesis of cognitive aging through reanalysis of Crawford et al.'s (2000) data. *Aging, Neuropsychology, and Cognition*, 9, 231–249. https://doi.org/10.1076/anec.9.3.231.9611

Flores, I., Casaletto, K. B., Marquine, M. J., Umlauf, A., Moore, D. J., Mungas, D., Gershon, R. C., Beaumont, J. L., & Heaton, R. K. (2017). Performance of Hispanics and Non-Hispanic whites on the NIH toolbox cognition battery: The roles of ethnicity and language backgrounds. *The Clinical Neuropsychologist*, 31, 783–797. https://doi.org/10.1080/13854046.2016.1276216

Gaudino, E. A., Geisler, M. W., & Squires, N. K. (1995). Construct validity in the trail making test: What makes part B harder? *Journal of Clinical and Experimental Neuropsychology*, 17, 529–535. https://doi.org/10.1080/01688639508405143

Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH toolbox for assessment of neurological and behavioral function. *Neurology*, 80, S2–S6. https://doi.org/10.1212/WNL.0b013e3182872e5f

Hackett, K., Krikorian, R., Giovannetti, T., Melendez-Cabrero, J., Rahman, A., Caesar, E. E., Chen, J. L., Hristov, H., Seifan, A., Mosconi, L., & Isaacson, R. S. (2018). Utility of the NIH toolbox for assessment of prodromal Alzheimer's disease and dementia. *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring*, 10, 764–772. https://doi.org/10.1016/j.dadm.2018.10.002

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S. N., & REDCap Consortium. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, 95, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42, 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Heaton, R. K., Akshoomoff, N., Tulsky, D., Mungas, D., Weintraub, S., Dikmen, S., Beaumont, J., Casaletto, K. B., Conway, K., Slotkin, J., & Gershon, R. (2014). Reliability and validity of composite scores from the NIH toolbox cognition battery in adults. *Journal of the International Neuropsychological Society*, 20, 588–598. https://doi.org/10.1017/S1355617714000241

Hessl, D., Sansone, S. M., Berry-Kravis, E., Riley, K., Widaman, K. F., Abbeduto, L., Schneider, A., Coleman. J., Oaklander, D., Rhodes, K. C., & Gershon, R. C. (2016). The NIH toolbox cognitive battery for intellectual disabilities: Three preliminary studies and future directions. *Journal of Neurodevelopmental Disorders*, 8, 35. https://doi.org/10.1186/s11689-016-9167-4

Hülür, G., Ram, N., Willis, S. L., Schaie, K. W., & Gerstorf, D. (2015). Cognitive dedifferentiation with increasing age and proximity of death: Within-person evidence from the Seattle longitudinal study. *Psychology and Aging*, 30, 311–323. https://doi.org/10.1037/a0039260

Kamin, S. T., & Lang, F. R. (2016). Cognitive functions buffer age differences in technology ownership. *Gerontology*, 62, 238–246. https://doi.org/10.1159/000437322

Lamar, M., Zonderman, A. B., & Resnick, S. (2002). Contribution of specific cognitive processes to executive functioning in an aging population. *Neuropsychology*, 16, 156–162. https://doi.org/10.1037/0894-4105.16.2.156

Lee Meeuw Kjoe, P. R., Agelink van Rentergem, J. A., Vermeulen, I. E., & Schagen, S. B. (2021). How to correct for computer experience in online cognitive testing? *Assessment*, 28, 1247–1255. https://doi.org/10.1177/1073191120911098

Ma, Y., Carlsson, C. M., Wahoske, M. L., Blazel, H. M., Chappell, R. J., Johnson, S. C., Asthana, S., & Gleason, C. E. (2021). Latent factor structure and measurement invariance of the NIH toolbox cognition battery in an Alzheimer's disease research sample. *Journal of the International Neuropsychological Society*, 27, 412–425. https://doi.org/10.1017/S1355617720000922

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163–203. https://doi.org/10.1037/0033-2909.109.2.163

MacLeod, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology: General*, 121, 12–14. https://doi.org/10.1037/0096-3445.121.1.12

Mungas, D., Heaton, R., Tulsky, D., Zelazo, P. D., Slotkin, J., Blitz, D., Lai, J. S., & Gershon, R. (2014). Factor structure, convergent validity, and discriminant validity of the NIH toolbox cognitive health battery (NIHTB-CHB) in adults. *Journal of the International Neuropsychological Society*, 20, 579–587. https://doi.org/10.1017/S1355617714000307

Mungas, D., Widaman, K., Zelazo, P. D., Tulsky, D., Heaton, R. K., Slotkin, J., Blitz, D.L., & Gershon, R. C. (2013). VII. NIH toolbox cognition battery (CB): Factor structure for 3 to 15 year olds. *Monographs of the Society for Research in Child Development*, 78, 103–118. https://doi.org/10.1111/mono.12037

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53, 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

O'Shea, D. M., Fieo, R., Woods, A., Williamson, J., Porges, E., & Cohen, R. (2018). Discrepancies between crystallized and fluid ability are associated with frequency of social and physical engagement in community dwelling older adults. *Journal of Clinical and Experimental Neuropsychology*, 40, 963–970. https://doi.org/10.1080/13803395.2018.1452195

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. https://doi.org/10.18637/jss.v048.i02

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, 132, 566–594. https://doi.org/10.1037/0096-3445.132.4.566

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10. https://doi.org/10.1007/BF02291170

Tulsky, D. S., Holdnack, J. A., Cohen, M. L., Heaton, R. K., Carlozzi, N. E., Wong, A. W. K., Boulton, A. J., & Heinemann, A. W. (2017). Factor structure of the NIH toolbox cognition battery in individuals with acquired brain injury. *Rehabilitation Psychology*, 62, 435–442. https://doi.org/10.1037/rep0000183

Victorson, D., Manly, J., Wallner-Allen, K., Fox, N., Purnell, C., Hendrie, H., Havlik, R., Harniss, M., Magasi, S., Correia, H., & Gershon, R. (2013). Using the NIH Toolbox in special populations: considerations for assessment of pediatric, geriatric, culturally diverse, non-English-speaking, and disabled individuals. *Neurology*, 80, S13–S19. https://doi.org/10.1212/WNL.0b013e3182872e26

Vincent, G. K., & Velkoff, V. A. (2010). *The next four decades the older population in the United States: 2010 to 2050*. U.S. Census Bureau.

Wallert, J., Rennie, A., Ferreira, D., Muehlboeck, J.-S., Wahlund, L.-O., Westman, E., Ekman, U., & ADNI consortium, and MemClin Steering Committee. (2021). Cognitive dedifferentiation as a function of cognitive impairment in the ADNI and MemClin cohorts. *Aging*, 13, 13430–13442. https://doi.org/10.18632/aging.203108

Weintraub, S., Bauer, P. J., Zelazo, P. D., Wallner-Allen, K., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Slotkin, J., Blitz, D. L., Carlozzi, N. E., Havlik, R. J., Beaumont, J. L., Mungas, D., Manly, J. J., Borosh, B. G., Nowinski, C. J., & Gershon, R. C. (2013). I. NIH toolbox cognition battery (CB): Introduction and pediatric data. *Monographs of the Society for Research in Child Development*, *78*, 1–15. https://doi.org/10.1111/mono.12031

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., & Gershon, R. C. (2013). Cognition assessment using the NIH toolbox. *Neurology*, *80*, S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Slotkin, J., Carlozzi, N. E., Bauer, P. J., Wallner-Allen, K., Fox, N., Havlik, R., Beaumont, J. L., Mungas, D., Manly, J. J., Moy, C., Conway, K., Edwards, E., Nowinski, C. J., & Gershon, R. (2014). 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*, *20*, 567–578. https://doi.org/10.1017/S1355617714000320

Werner, J. M., Carlson, M., Jordan-Marsh, M., & Clark, F. (2011). Predictors of computer use in community-dwelling, ethnically diverse older adults. *Human Factors*, *53*, 431–447. https://doi.org/10.1177/0018720811420840

Wilson, R. S., Segawa, E., Hizel, L. P., Boyle, P. A., & Bennett, D. A. (2012). Terminal dedifferentiation of cognitive abilities. *Neurology*, *78*, 1116–1122. https://doi.org/10.1212/WNL.0b013e31824f7ff2

Wu, Y.-H., Lewis, M., & Rigaud, A.-S. (2019). Cognitive function and digital device use in older adults attending a memory clinic. *Gerontology & Geriatric Medicine*, *5*, 2333721419844886. https://doi.org/10.1177/2333721419844886