Validity and Normative Data for the Biber Figure Learning Test: A Visual Supraspan Memory Measure

Katherine A. Gifford, PsyD,[a] Dandan Liu, PhD,[b] Jacquelyn E. Neal, BA,[b] Lealani Mae Y. Acosta, MD, MPH[a], Susan P. Bell, MBBS, MSCI[a,c], Margaret E. Wiggins, BA,[d] Kristi M. Wisniewski, BS,[e] Mary Godfrey, BA,[f] Laura A. Logan, BS,[a] Timothy J. Hohman, PhD,[a] Kimberly A. Pechman, PhD,[a] David J. Libon, PhD,[g] Kaj Blennow, MD, PhD, [h,i] Henrik Zetterberg, MD, PhD,[h,i,j,k] & Angela L. Jefferson, PhD[a]

[a]*Vanderbilt Memory & Alzheimer's Center, Department of Neurology, Vanderbilt University Medical Center, Nashville, TN;*
[b]*Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN;*
[c]*Divisions of Cardiovascular and Geriatric Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA*
[d]*Department of Clinical and Health Psychology, University of Florida, Gainesville, FL*
[e]*School of Gerontology, University of Southern California, Los Angeles, CA*
[f]*Department of Psychology, Drexel University, Philadelphia, PA*
[g]*Department of Geriatrics and Gerontology, Rowan University – School of Osteopathic Medicine, Stratford, New Jersey*
[h]*Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, The Sahlgrenska Academy at University of Gothenburg, Mölndal, Sweden*
[i]*Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden*
[j]*Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, United Kingdom*
[k]*UK Dementia Research Institute at UCL, London, United KIngdom*

Corresponding Author:
Angela L. Jefferson, PhD
Vanderbilt University Medical Center
Department of Neurology
Vanderbilt Memory & Alzheimer's Center
1207 17th Avenue S., Suite 204
Nashville, TN 37212
Phone: 615-322-8676
Email: angela.jefferson@vanderbilt.edu

Acknowledgements and Funding: This work was supported by K24-AG046373 (ALJ); Alzheimer's Association IIRG-08-88733 (ALJ); R01-AG034962 (ALJ); R01-HL111516 (ALJ); R01-NS100980 (ALJ); K12-HD043483 (KAG/TJH/SPB); Alzheimer's Association NIRG-13-283276 (KAG); K23-045966 (KAG); Paul B. Beeson Career Development Award in Aging K23-AG048347 (SPB), the Eisenstein Women's Heart Fund (SBP),

K01-AG049164 (TJH), UL1-TR000445 (Vanderbilt Institute for Clinical & Translational Research); the Vanderbilt Memory & Alzheimer's Center.

Abstract

This study evaluated the biological correlates and psychometrics of the Biber Figure Learning Test (BFLT), a visuospatial serial figure learning test, for use in detecting cognitive impairment in older adults. Non-demented individuals (n=322, 72±7, 41% female) from the Vanderbilt Memory & Aging Project completed a comprehensive neuropsychological protocol. Adjusted regression models related BFLT indices to structural brain MRI and cerebrospinal fluid markers (CSF) of brain health. Psychometric properties were examined, including construct validity and regression-based normative data. Lower BFLT performance related to smaller medial temporal lobe volumes and higher CSF tau concentrations. BFLT indices were most strongly correlated with other measures of verbal and non-verbal memory measures and executive functioning. The BFLT comprehensively assesses all aspects of visuospatial learning and memory and is sensitive to markers of unhealthy brain aging. Enhanced normative data enriches the clinical utility of this visual serial figure-learning test for use with older adults.

Key words: episodic memory, visual memory, normative data, aging, regression-based norms

Introduction

Alzheimer's disease (AD) is a major public health issue, and early diagnosis is critical to managing disease burden. As such, neuropsychological assessment, specifically episodic learning and memory, is an integral component for early and accurate diagnosis of AD (Karantzoulis & Galvin, 2011). Clinically, AD is often characterized by an insidious decline in episodic memory, particularly in the early phases of illness. Serial verbal list-learning tests (e.g., Brandt, 2001; Delis, Kramer, Kaplan, & Ober, 2000) have demonstrated great utility in early detection of AD because they assess key constructs underlying memory and learning, including encoding or rate of acquisition, retrieval, and susceptibility to proactive and retroactive interference (Albert et al., 2011). Examples of such verbal supraspan tests include the California Verbal Learning Test-Second Edition (CVLT-II; Delis et al., 2000), the Rey Auditory Verbal Learning Test (Tierney et al., 1994), and the Hopkins Verbal Learning Test-Revised (Brandt, 2001).

Surprisingly, a non-verbal analogue to these verbal list-learning tests is lacking given the number of commonly used tools to assess aspects of visual encoding, retrieval, and memory. While each measure is valuable in assessing some aspects of visual learning and memory, no single tool reflects the full range of episodic memory constructs that are essential to define in clinical practice. For example, common visual measures, such as Rey-Osterrieth Complex Figure Test (Osterrieth, 1944; Rey, 1941), WMS Visual Reproduction (Wechsler, 2009), or Benton Visual Retention Test (BVRT; Sivan, 1992), assess memory using a single exposure or learning trial, limiting measurement of learning (acquisition). While tests such as the Brief Visuospatial

Memory Test-Revised (Benedict, Schretlen, Groninger, Dobraski, & Shpritz, 1996) and the Visual Spatial Learning Test (Malec, Ivnik, & Hinkeldey, 1991) include presentation of material over several learning trials, neither tool assesses vulnerability to proactive or retroactive interference through introduction of a distractor trial. Thus, while several visual learning and memory tools are available for clinical implementation, their varied methodologies limit more detailed assessment of well-known constructs underlying episodic memory, including encoding, retrieval, recognition, and susceptibility to interference. Further, the existing paradigms preclude direct comparison with existing verbal supraspan tests.

The Biber Figure Learning Test (BFLT) is a visuospatial serial figure learning test designed to assess key components underlying episodic memory (Glosser, Goodglass, & Biber, 1989). It was modelled after supraspan tests (Delis, Kramer, Kaplan, & Ober, 1987; Delis et al., 2000) and Rey's visuospatial task (Rey, 1968), and was originally developed as a 10-item memory test (Glosser et al., 1989). The BFLT (extended; Glosser, Cole, Khatri, DellaPietra, & Kaplan, 2002) is comprised of five immediate free recall learning trials of 15 non-familiar geometric shapes of moderate visual complexity, an interference condition that presents a new list of 15 items (Distractor Trial), short and long delay free recall trials for the original 15 items, and a recognition trial that includes the original items, interference items, and new test items.

The BFLT offers several advantages over other available visual memory tools. First, it is a visuospatial analogue to verbal list-learning tasks with indices capturing encoding/rate of acquisition, retrieval, susceptibility to proactive and retroactive interference, and recognition test performance. Thus, when used together, both tools

provide parallel information to inform the potential presence of modality-specific memory impairment. Second, the BFLT has an alternate test form for repeat administration needs. Third, the BFLT includes an abbreviated 10-item version to accommodate patients suspected of more compromised neuropsychological function, such as dementia (Glosser et al., 2002; Glosser et al., 1989). Collectively, the BFLT provides an opportunity to comprehensively assess specific visual episodic modalities using a format consistent with a verbal serial list-learning paradigm, making the BFLT potentially more useful than many widely used visual memory tools.

Despite the BFLT's potential clinical utility, the test has not been widely used. Foremost, the validity of this visuospatial memory test is not well delineated. Establishing a link between BFLT indices and markers of unhealthy brain aging, including AD, such as amyloid, neurodegeneration, and white matter disease is essential to demonstrating the utility and validity of this test in older adult populations. Existing BFLT data have been restricted to individuals with epilepsy or non-AD amnestic syndromes implicating medial temporal lobe and subcortical-limbic pathways (Glosser et al., 1989), particularly within the right hemisphere (Glosser et al., 2002). Similarly, establishing evidence of convergent and discriminant validity with other neuropsychological measures is needed. Second, there is a lack of comprehensive normative data; existing normative data is restricted to a few dozen individuals and upper age bands (e.g., 40 to 79 years; Glosser et al., 1989) with minimal consideration of how important demographic variables, such as age, sex, education, and race/ethnicity (Norman, Evans, Miller, & Heaton, 2000) may confound performance.

Leveraging a community-based sample, this study has several objectives. First, we aim to examine the biological correlates of the BFLT, hypothesizing that BFLT indices would relate to markers of unhealthy brain aging including amyloid pathology (cerebrospinal fluid (CSF) amyloid beta42; Mandecka et al., 2016), neurodegeneration (hippocampal volume on brain MRI, inferior lateral ventricle volume on brain MRI, and CSF tau; Mandecka et al., 2016; Murphy et al., 2010; Wolk & Dickerson, 2011) and white matter disease (white matter hyperintensities (WMH) on brain MRI and CSF neurofilament light; Lee et al., 2016; Zetterberg et al., 2016). Second, we will test associations between the BFLT and other neuropsychological tests, hypothesizing BFLT indices will correlate most strongly with tests assessing verbal learning and memory (Benedict et al., 1996) and visuospatial and executive function (Jefferson et al., 2006) but not language (Meyers & Meyers, 1995). Lastly, we provide regression-based normative data accounting for demographic variables that often confound cognitive performance including age, sex, race/ethnicity, and education (Glosser et al., 1989; Norman et al., 2000). These efforts will enhance the BFLT's utilization as a more valuable tool in the assessment of visuospatial serial figure learning task in older adults.

Methods

Participants

Participant data were drawn from the Vanderbilt Memory & Aging Project (Jefferson et al., 2016), a longitudinal observational study investigating vascular health and unhealthy brain aging. Participants were recruited through postal mailings, radio advertisements, newsletters, research distribution emails, community events, websites, and word-of-mouth. Inclusion required participants be age 60 years or older, speak English, have adequate auditory and visual acuity for testing purposes, and have a reliable study partner defined as someone the participant has known for a minimum of two years, with weekly contact, and knowledge of the participant's cognitive and functional abilities. At eligibility, participants underwent medical history and record review, clinical interview with the participant and their informant using the Clinical Dementia Rating (CDR; Morris, 1993) and Functional Activities Questionnaire (Pfeffer, Kurosaki, Harrah, Chance, & Filos, 1982), and a comprehensive neuropsychological protocol assessing multiple cognitive systems. Collectively, this information was used by a consensus team to determine each participant's cognitive diagnosis, including:

1. Cognitively normal participants (NC): defined as (a) CDR=0 (no dementia), (b) no deficits in activities of daily living directly attributable to cognitive impairment, and (c) no evidence of neuropsychological impairment defined as standard scores falling within 1.5 standard deviations of the age-adjusted normative mean.
2. Early MCI: defined as (a) Clinical Dementia Rating (CDR)=0.5 (reflecting mild severity of impairment), (b) no deficits in activities of daily living attributable to

cognitive issues, and (c) no neuropsychological impairment defined as standard scores falling within 1.5 standard deviations of the age-adjusted normative mean (Aisen et al., 2010).

3. MCI: defined as (a) CDR=0 or 0.5 (reflecting mild severity of impairment), (b) relatively spared activities of daily living, (c) neuropsychological impairment within at least one cognitive domain (i.e., performances falling greater than 1.5 standard deviations outside the age-adjusted normative mean or pre-morbid level of functioning), (d) concern of a cognitive change by the participant, informant, or clinician, and (e) absence of a dementing syndrome (Albert et al., 2011).

Participants were excluded for a diagnosis other than normal cognition, early MCI, or MCI, history of neurological disease (e.g., dementia, multiple sclerosis, stroke), heart failure, major psychiatric illness, head injury with loss of consciousness >5 minutes, and systemic or terminal illness that could affect the individual's ability to participate in follow-up examinations. Ultimately, this study enrolled 335 community-dwelling individuals age 60 to 92 years.

Neuropsychological Assessment

As part of the study enrollment visit, participants completed a neuropsychological protocol assessing multiple cognitive systems. Note, this protocol was different than what was used for eligibility determination and cognitive diagnosis at study entry. The BFLT methods are described in the introduction above and indices analyzed in the current study include: Trial 1 Learning, Trial 2 Learning, Trial 3 Learning, Trial 4

Learning, Trial 5 Learning, Trial 1-5 Total Learning, Distractor Trial (Trial B), Short Delay Free Recall, Long Delay Free Recall (following a 20 minute filled delay), Long Delay Recognition Measures (including Recognition Total Correct, Distractor Trial False Alarms, Novel False Alarms, Total False Alarms, Long Delay Recognition Total Discrimination calculated as follows: ((Recognition Total Correct + 0.5)/16) – ((Total False Alarms + 0.5)/31)), Total Repetitions, and Total Intrusions. See **Table 1** for a more detailed description of these indices and information regarding the remaining measures in the neuropsychological protocol.

Brain MRI

Participants were scanned at the Vanderbilt University Institute of Imaging Science on a 3T Philips Achieva system (Best, The Netherlands) with 8-channel SENSE reception (Pruessmann, Weiger, Scheidegger, & Boesiger, 1999). T_1-weighted MPRAGE (isotropic spatial resolution=1mm) images were post-processed for tissue volume quantification using an established Multi-Atlas Segmentation (Asman & Landman, 2012). Specific regions of interest included hippocampal volume and inferior lateral ventricle volume measured within each hemisphere and then summed across hemispheres for a total volume. Intracranial volume was defined using voxel-based morphometry in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). T_2-weighted fluid-attenuated inversion recovery (FLAIR) was acquired for quantification of WMH and post-processed using the Lesion Segmentation Tool toolbox for SPM8 (Schmidt et al., 2012).

Cerebrospinal Fluid Acquisition

A subset of individuals (82 participants) completed an optional morning fasting lumbar puncture. CSF was collected with polypropylene syringes using a Sprotte 25-gauge spinal needle in the L3/L4 or L4/L5 intervertebral lumbar space. Samples were immediately mixed and centrifuged, and supernatants were aliquoted in 0.5mL polypropylene tubes and stored at -80°C. Samples were analyzed in batch using commercially available enzyme-linked immunosorbent assays (Fujirebio, Ghent, Belgium) to determine the levels of Aβ1-42 (INNOTEST® β-AMYLOID$_{(1-42)}$), total tau (INNOTEST® hTAU), and tau phosphorylated at threonine 181 (p-tau; INNOTEST® PHOSPHO-TAU$_{(181P)}$). Neurofilament Light (NFL) was measured using a commercially available ELISA (Uman Diagnostics). Processing was completed by board certified laboratory technicians who were blinded to clinical information (Palmqvist et al., 2014). Intra-assay CVs were <10%.

Statistical Analysis

Descriptive statistics for participant characteristics including age, sex, race/ethnicity, education, mood (assessed with Geriatric Depression Scale (GDS) total score; Yesavage et al., 1983), CSF levels, brain MRI variables, and neuropsychological performances (see **Table 1** for a comprehensive list of measures) were calculated. To assess construct validity, linear regression related selected BFLT indices (Total Learning, Short Delay Recall, Long Delay Recall, Total Discrimination) to markers of unhealthy brain aging adjusting for age, sex, education, race/ethnicity, and diagnosis in the entire cohort. Specifically, we assessed amyloidosis (CSF AB_{42}), neurodegeneration

(CSF tau, CSF p-tau, hippocampal volume, inferior lateral ventricle volume), and white matter disease (WMH, CSF NFL). WMH values were log-transformed due to non-normal distributions. Models analyzing brain volume outcomes additionally adjusted for intracranial volume. Convergent and discriminant validity were assessed in the NC subsample using Spearman's rank correlation coefficients to relate selected BFLT indices with other neuropsychological variables and GDS total score (see **Table 1** for a comprehensive list of measures). Regression-based normative data of the above selected BFLT indices was generated from the NC subgroup using multiple regression analyses adjusting for age, sex, education, and race/ethnicity. The raw score on the selected BFLT indices were used as outcomes. Sex was coded as male=0, female=1 and race/ethnicity was coded as White/Non-Hispanic=0, Non-White/Hispanic=1. Age and education in years were treated as continuous variables. Multi-collinearity was assessed by calculating variance inflation factors (VIFs) and residuals plots were visually inspected for goodness of fit. Intercepts, beta-coefficients, and the root-mean-squared error for each model were presented for calculation of a predicted BFLT performance value using the following equation (see Equation 1):

(1) Predicted score= β_0 (intercept) + β_{age} X age (actual age in years) + β_{sex} X sex (0=male, 1=female) + $\beta_{race/ethnicity}$ X race/ethnicity (0=White/Non-Hispanic, 1=Non-White/Hispanic) + $\beta_{education}$ X education (years of education completed).

To calculate a z-score normative value, the predicted score calculated in Equation 1 was used within the following equation (see Equation 2; Shirk et al., 2011):

(2) z-score = (observed score – predicted score)/root-mean-squared-error.

Significance was set *a priori* at $p<0.05$. All analyses were conducted using *R 3.2.0* (www.r-project.org).

Results

Participant Characteristics

The entire cohort comprised 322 participants with an age range of 60 to 92 years (73±7) and an education range of 7 to 20 years (16±3). Approximately half (51%, n=174) of individuals had normal cognition, 41% were women and 87% self-declared as White/Non-Hispanic. Global cognition, as assessed by the Montreal Cognitive Assessment (MoCA), ranged 14 to 30 (25.4±3.3). **Table 2** contains full details on participant characteristics, brain volumes, CSF levels, and neuropsychological performances.

Normative data analyses included only NC participants, representing a subsample of 174 individuals with an age range of 60 to 92 years (72±7) and 41% women. A majority of NC participants self-declared as White/Non-Hispanic (87%) with an education range of 10 to 20 years (16±3). MoCA scores ranged from 17 to 30 (27.0±2.2). Please see **Table 2** for full details on NC participant characteristics and **Table 3** for neuropsychological performances.

BFLT & Biomarkers of Brain Health

Amyloidosis: None of the examined BFLT indices were associated with CSF $A\beta_{42}$ values (all p-values>0.05; see **Table 4**).

Neurodegeneration: Worse performance on all BFLT indices related to CSF markers of neurodegeneration, including higher t-tau (p-values<0.01) and p-tau levels (p-values<0.02). Similarly, lower scores on all BFLT variables related to smaller total hippocampal volume (p-values<0.01), smaller left hippocampal volume (p-

values<0.001), larger total inferior lateral ventricle volume (p-values<0.05), and larger left inferior lateral ventricle volume (p-values<0.03). Only BFLT Short Delay Free Recall and Long Delay Free Recall were positively associated with right hippocampal volume (p-values<0.03). Only BFLT Long Delay Free Recall and Long Delay Total Discrimination were negatively associated with right inferior lateral ventricle volume (p-values<0.05; see **Table 4**).

White Matter Disease: None of the BFLT indices were related to WMH volume (all p-values>0.12; see **Table 4**).

BFLT Correlations with Other Cognitive Measures

Global Cognition: MoCA performance was moderately correlated with all BFLT indices (all p-values<0.001, see **Table 5**).

Verbal Learning & Memory: CVLT-II Total Learning, Short Delay Free Recall, Long Delay Free Recall, and Recognition Total Discrimination correlated with all selected BFLT indices (p-values<0.001; see **Table 5**). CVLT-II Distractor Trial was correlated with most BFLT indices (p<0.02; see **Table 5**) but was unrelated to BFLT Recognition Total Discrimination (p=0.09; see **Table 5**)**.**

Visuospatial Learning & Memory: BVRT Total Score correlated with all selected BFLT indices (p-values<0.002; see **Table 5**).

Visuospatial Abilities: Wechsler Adult Intelligence Scale-IV (WAIS-IV; Wechsler, 2008) Block Design correlated with most BFLT indices (p-values<0.001) but not BFLT Distractor Trial (Trial B; p=0.08; see **Table 5**). Hooper Visual Organization Test (Hooper, 1983) performance was related to all selected BFLT indices (p-values<0.001).

See **Table 5** for full details.

Language: Both Boston Naming Test (30-item) performance (p-values<0.006) and Category Fluency performance (p-values<0.004) were correlated with all BFLT indices. See **Table 5** for full details.

Attention/Information Processing Speed: Delis Kaplan Executive Functioning System (DKEFS; Delis, Kaplan, & Kramer, 2001) Number Sequencing Test was related to most BFLT indices (p-values<0.001) except for BFLT Distractor Trial (p=0.05; see **Table 5** for details). WAIS-IV Coding performance related to all BFLT indices (all p-values<0.001; see **Table 5** for details).

Executive Functioning: DKEFS Tower Test performance was related to all BFLT indices (p-values<0.03; see **Table 5**). DKEFS Color-Word Interference Test Inhibition performance was related to most BFLT indices (p-values<0.02) except BFLT Recognition Total Discrimination (p=0.09; **Table 5**). DKEFS Number-Letter Switching Test and Letter Fluency were correlated with all BFLT indices (all p-values<0.001). See **Table 5** for full details.

Mood: GDS was unrelated to all BFLT indices (all p-values>0.37). See **Table 5** for details.

BFLT Regression-Based Normative Data

Review of VIF revealed no multicollinearity between predictor and demographic variables (all VIF<1.11). The residual plots against predicted values did not reveal any systematic patterns, suggesting sufficient goodness of fit. Means, intercepts, and regression coefficients are presented in **Table 6** for transforming raw scores into

demographically adjusted z-scores using Equations 1 and 2 in the Statistical Analysis section.

For illustrative purposes, normative data for BFLT Total Learning is calculated for an individual with the following demographics: 75 year old female, White/Non-Hispanic, with 16 years of education with a BFLT Total Learning score of 130. Using Equation 1, the predicted score is calculated as follows:

223.7 (BFLT Total Learning intercept) + -1.79x75 (BFLT Total Learning β_{age} X actual age) + 12.78x1 (BFLT Total Learning β_{sex} X 0=male) + -10.30x0 (BFLT Total Learning $\beta_{race/ethnicity}$ X 0=White/non-Hispanic) + 2.26x16 (BFLT Total Learning $\beta_{education}$ X # years of education completed) = predicted score of 138.39.

To calculate a normative value with an obtained score of 130, Equation 2 is used:

130 (observed score) – 138.39 (predicted score)/27.25 (root-mean-squared-error for BFLT Total Learning) = z-score of -0.31.

Discussion

The BFLT is a visuospatial serial figure-learning test that allows for comprehensive assessment of visual learning and memory and offers advantages to existing visual memory tests. This study aimed to improve the clinical utility of the BFLT among older adults by demonstrating the test's psychometric properties, including validity. Specifically, BFLT performances were broadly related to CSF tau and medial temporal lobe brain volumes, but minimal associations with CSF amyloid or markers of white matter disease (WMH or CSF NFL) were observed. Furthermore, compared to other neuropsychological tests, BFLT indices correlated with measures of verbal learning and memory, information processing, executive functioning, and visuospatial performances, as expected. BFLT was also (unexpectedly) related to language tests. To further maximize the measure's usefulness, regression-based normative data is provided with adjustments for common demographic confounds.

The current results suggest that performance on the BFLT is related to medial temporal lobe structures involved in successful learning and memory. These structures include the hippocampus and surrounding perirhinal areas critical for the formation of new memories (Fell et al., 2001). These regions have been linked with performance on verbal supraspan episodic memory tests (Wolk & Dickerson, 2011) and existing visual memory tests (Buffalo, Reber, & Squire, 1998). The current results are among the first to highlight the construct validity of the BFLT to measure integrity of learning and memory structures.

In addition to structural evidence of neurodegeneration, the BFLT was associated with CSF tau, another marker of neurodegeneration, but not amyloidosis or white matter

disease. Early patterns of atrophy seen on brain MRI are most notable within the medial temporal lobe (Convit et al., 2000), including the hippocampus (Fox et al., 1996) and inferior lateral ventricles (Thompson et al., 2004) because this region is the first to be affected by AD pathology. Tau deposition begins in the entorhinal cortex, spreading to the hippocampus and surrounding medial temporal regions (Braak & Braak, 1991; Duyckaerts, Delatour, & Potier, 2009), resulting in axonal loss and neurodegeneration. Tau deposition levels and brain volume loss are thought to be closely related (Fjell et al., 2010) and represent disease or symptom severity (Blennow, 2004). Additionally, tau correlates with objective cognitive performance (Samgard et al., 2010), including episodic memory (Fjell et al., 2008), often more closely than markers of amyloidosis (Brier et al., 2016). Taken cumulatively, the BFLT may be an important tool to assess for underlying neurodegenerative disorders.

The stronger association between BFLT indices and regions within the left medial temporal lobe as compared to the right was somewhat unexpected given earlier work linking the right medial temporal lobe with BFLT performances in epilepsy or non-AD related amnestic syndromes (Glosser et al., 2002; Glosser et al., 1989). However, this discrepant finding may be related to differing methodology given the previous results did not include any neuroimaging analyses and focused on clinical presentation of epilepsy patients. The current finding highlights the importance of left medial temporal lobe integrity for visual memory tasks, an observation supported by a functional MRI study linking activation in the left hippocampus to visual memory (Ranganath, Cohen, Dam, & D'Esposito, 2004). Alternatively, the right hippocampus may be integral in encoding/recalling information regarding spatial navigation or location (Burgess,

Maguire, & O'Keefe, 2002; Kesner, Bolland, & Dakis, 1993), a component of visuospatial memory and processing that is less essential to successful BFLT performance. Notably, the current findings suggest bilateral hippocampal integrity for retrieval of visual information, consistent with a recent meta-analysis implicating both the right and left hippocampus in retrieval of visual information (Lepage, Habib, & Tulving, 1998). The preferential association with the left hemisphere in encoding visual information provides additional evidence that individuals often utilize verbal strategies to encode visual material, even non-meaningful visual information, thus employing left hemisphere regions.

Surprisingly, there was a paucity of associations between BFLT indices and white matter disease. These null associations were surprising given the link between white matter disease and verbal memory, including supraspan paradigms analogous to the BFLT (Kennedy & Raz, 2009). Also, BFLT performance was related to measures of information processing speed and executive functioning, cognitive domains that are often closely correlated with WMH burden (Madden, Bennett, & Song, 2009; Prins et al., 2005). The null findings could be due to the inclusion of cognitively normal participants in the analyses as well as the low prevalence of cardiovascular disease across the entire sample.

BFLT related to verbal memory and visuospatial performances as measured by other popular tools (Benton, 1974; Delis et al., 2000). Similarly, BFLT correlated with executive function performances, an expected finding given the executive demands of the BFLT and existing support linking visuospatial abilities with executive skills. Prior research has suggested that executive functions, such as the capability to monitor and

switch between the target and distractor set of stimuli and the ability to organize the geometric shapes into accurate designs, is important during visuospatial tasks (Jefferson et al., 2006; Libon et al., 1994). Unexpectedly, BFLT was related to language performances possibly because individuals with more intact language abilities utilize verbal cues or assign common object descriptions to the BFLT design stimuli as a tool for successfully remembering design details (Schacter, Cooper, & Delaney, 1990). This idea is supported by our finding that BFLT performance appeared preferentially related to left hemisphere medial temporal lobe functioning and suggesting that successful performance on BFLT indices relies on integrity of brain regions also important for language abilities.

Advancing age was noted to be related to poorer performance on all BFLT indices, consistent with previously research linking increasing age with declines in verbal memory (Gunning-Dixon & Raz, 2003; Norman et al., 2000) and visual retention (Coman et al., 1999). Sex, race/ethnicity, and education were inconsistently related to the various BFLT indices, although results are aligned with previous research suggesting better cognitive performance is related to more years of education, female sex, and White/non-Hispanic ethnicity. Given these potential demographic confounds on task performance, the regression-based normative data provided here incorporate all of these demographic factors, allowing for more robust generation of normative data. Additionally, the sample size of 174 well-characterized, cognitively normal, adults age 60 to 92 represents an improvement to the existing normative data, facilitating more widespread clinical use of the BFLT.

The current study has several strengths. First, the extensive phenotyping of cognitive status for all enrolled participants included a CDR interview with the participant and a reliable informant, medical record and health history review, comprehensive neuropsychological protocol, and consensus decision for diagnostic status by experienced clinicians. Second, the comprehensive neuropsychological protocol examined in the current methods (which was separate from the protocol used to diagnose and enroll participants as NC) encompassed multiple cognitive domains, permitting detailed comparisons between BFLT and other cognitive task performances. Third, the normative sample size of the current study (n=174) and use of regression norms considering multiple demographic confounders enhances the one existing BFLT normative data report (Glosser et al., 2002). Fourth, the correlation data with other common neuropsychological tests, including an analogous verbal supraspan test, also provide context for interpreting underlying BFLT impairments in a comprehensive clinical evaluation. Lastly, this study is the first linking the BFLT to biological markers of brain health to assess the measure's validity.

Despite these strengths, there are several limitations. First, aspects of the cohort could limit the generalizability to the findings. For example, the sample is predominantly White with a mean college education level. Similarly, the inclusion and exclusion criteria associated with enrollment into the Vanderbilt Memory & Aging Project, including minimal cardiovascular disease, may limit generalized applicability. Second, the current study includes a smaller sample size in comparison to previous normative research with other learning and memory tests (Fine, Kramer, Lui, Yaffe, & Study Of Osteoporotic Fractures Sof Research, 2012; Gallassi et al., 2014). However, regression-based

normative procedures require smaller samples sizes than traditional-based normative procedures (Oosterhuis, van der Ark, & Sijtsma, 2016), and this study represents an initial and more comprehensive approach.

Overall, our findings provide novel information about the BFLT to enhance the clinical utility of this visuospatial serial figure learning test. The BFLT measures brain health and integrity through the assessment of multiple aspects of learning and memory, including learning or rate of acquisition, encoding or storage, retrieval, recognition, and freedom from interference. Results suggest the BFLT has good psychometric properties and relates to medial temporal lobe integrity. Regression-based normative data based upon age, sex, education, and race/ethnicity are provided to enhance clinical utility of this promising visuospatial memory test. Further work is needed linking BFLT performance to longitudinal change in markers of pathology.

Disclosure Statement: No relevant financial or nonfinancial relationships exist for any authors**.**

References

Aisen, P. S., Petersen, R. C., Donohue, M. C., Gamst, A., Raman, R., Thomas, R. G., . . . Alzheimer's Disease Neuroimaging, I. (2010). Clinical Core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans. *Alzheimers Dement, 6*(3), 239-246. doi:10.1016/j.jalz.2010.03.006

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 270-279. doi:10.1016/j.jalz.2011.03.008

Asman, A. J., & Landman, B. A. (2012). Formulating spatially varying performance in the statistical fusion framework. *IEEE Transactions on Medical Imaging, 31*(6), 1326-1336. doi:10.1109/TMI.2012.2190992

Benedict, R. H., Schretlen, D., Groninger, L., Dobraski, M., & Shpritz, B. (1996). Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability, and validity. *Psychological Assessment, 8*(2), 145-153.

Benton, A. L. (1974). *Revised Visual Retention Test: Clinical and experimental applications (4th edition)*. New York: The Psychological Corporation.

Benton, A. L., Hamsher, K., & Sivan, A. B. (1994). *Multilingual Aphasia Examination (3rd ed.)*. Iowa City: AJA Associates.

Blennow, K. (2004). CSF biomarkers for mild cognitive impairment. *Journal of Internal Medicine, 256*(3), 224-234.

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica, 82*(4), 239-259.

Brandt, J. B., R. (2001). *Hopkins Verbal Learning Test-Revised: Professional Manual*. Florida: PAR.

Brier, M. R., Gordon, B., Friedrichsen, K., McCarthy, J., Stern, A., Christensen, J., . . . Ances, B. M. (2016). Tau and Abeta imaging, CSF measures, and cognition in Alzheimer's disease. *Sci Transl Med, 8*(338), 338ra366. doi:10.1126/scitranslmed.aaf2362

Buffalo, E. A., Reber, P. J., & Squire, L. R. (1998). The human perirhinal cortex and recognition memory. *Hippocampus, 8*(4), 330-339. doi:10.1002/(SICI)1098-1063(1998)8:4<330::AID-HIPO3>3.0.CO;2-L

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron, 35*(4), 625-641.

Coman, E., Moses, J. A., Jr., Kraemer, H. C., Friedman, L., Benton, A. L., & Yesavage, J. (1999). Geriatric performance on the Benton Visual Retention Test: demographic and diagnostic considerations. *Clin Neuropsychol, 13*(1), 66-77. doi:10.1076/clin.13.1.66.1972

Convit, A., de Asis, J., de Leon, M. J., Tarshish, C. Y., De Santi, S., & Rusinek, H. (2000). Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. *Neurobiology of Aging, 21*(1), 19-26.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS): Examiner's Manual*. San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Adult Version*. San Antonio, TX: The Psychological Corporation.

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

Duyckaerts, C., Delatour, B., & Potier, M. C. (2009). Classification and basic pathology of Alzheimer disease. *Acta Neuropathol, 118*(1), 5-36. doi:10.1007/s00401-009-0532-1

Fell, J., Klaver, P., Lehnertz, K., Grunwald, T., Schaller, C., Elger, C. E., & Fernandez, G. (2001). Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nat Neurosci, 4*(12), 1259-1264. doi:10.1038/nn759

Fine, E. M., Kramer, J. H., Lui, L. Y., Yaffe, K., & Study Of Osteoporotic Fractures Sof Research, G. (2012). Normative data in women aged 85 and older: verbal fluency, digit span, and the CVLT-II short form. *The Clinical Neuropsycholist, 26*(1), 18-30. doi:10.1080/13854046.2011.639310

Fjell, A. M., Walhovd, K. B., Amlien, I., Bjornerud, A., Reinvang, I., Gjerstad, L., . . . Fladby, T. (2008). Morphometric changes in the episodic memory network and tau pathologic features correlate with memory performance in patients with mild cognitive impairment. *AJNR Am J Neuroradiol, 29*(6), 1183-1189. doi:10.3174/ajnr.A1059

Fjell, A. M., Walhovd, K. B., Fennema-Notestine, C., McEvoy, L. K., Hagler, D. J., Holland, D., . . . Dale, A. M. (2010). CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer's disease. *Journal of Neuroscience, 30*(6), 2088-2101. doi:30/6/2088 [pii]

10.1523/JNEUROSCI.3785-09.2010

Fox, N. C., Warrington, E. K., Freeborough, P. A., Hartikainen, P., Kennedy, A. M., Stevens, J. M., & Rossor, M. N. (1996). Presymptomatic hippocampal atrophy in Alzheimer's disease. A longitudinal MRI study. *Brain, 119 (Pt 6)*, 2001-2007.

Gallassi, R., Sambati, L., Stanzani Maserati, M., Poda, R., Oppi, F., De Matteis, M., & Marano, G. (2014). Simple verbal analogies test: normative data on a short task exploring abstract thinking. *Aging Clin Exp Res, 26*(1), 67-71. doi:10.1007/s40520-013-0180-0

Glosser, G., Cole, L., Khatri, U., DellaPietra, L., & Kaplan, E. (2002). Assessing nonverbal memory with the Biber Figure Learning Test - extended in temporal lobe epilepsy patients. *Archives of Clinical Neuropsychology, 17*(1), 25-35.

Glosser, G., Goodglass, H., & Biber, C. (1989). Assessing visual memory disorders. *Journal of Consulting and Clinical Psychology, 1*, 82-91.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia, PA: Lea & Febiger.

Gunning-Dixon, F. M., & Raz, N. (2003). Neuroanatomical correlates of selected executive functions in middle-aged and older adults: a prospective MRI study. *Neuropsychologia, 41*(14), 1929-1941.

Hooper, H. (1983). *Hooper Visual Organization Test (HVOT)*. Los Angeles, CA: Western Psychological Services.

Jefferson, A. L., Gifford, K. A., Acosta, L. M., Bell, S. P., Donahue, M. J., Taylor Davis, L., . . . Liu, D. (2016). The Vanderbilt Memory & Aging Project: Study Design and Baseline Cohort Overview. *Journal of Alzheimer's Disease, 52*(2), 539-559. doi:10.3233/JAD-150914

Jefferson, A. L., Wong, S., Bolen, E., Ozonoff, A., Green, R. C., & Stern, R. A. (2006). Cognitive predictors of HVOT performance differ between patients with mild cognitive impairment and normal controls. *Archives of Clinical Neuropsychology, 21*(8), 752-754.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test* (2nd ed.). Philadelphia, PA: Lea & Febiger.

Karantzoulis, S., & Galvin, J. E. (2011). Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Review of Neurotherapeutics, 11*(11), 1579-1591. doi:10.1586/ern.11.155

Kennedy, K. M., & Raz, N. (2009). Aging white matter and cognition: differential effects of regional variations in diffusion properties on memory, executive functions, and speed. *Neuropsychologia, 47*(3), 916-927. doi:10.1016/j.neuropsychologia.2009.01.001

Kesner, R. P., Bolland, B. L., & Dakis, M. (1993). Memory for spatial locations, motor responses, and objects: triple dissociation among the hippocampus, caudate nucleus, and extrastriate visual cortex. *Exp Brain Res, 93*(3), 462-470.

Lee, S., Viqar, F., Zimmerman, M. E., Narkhede, A., Tosto, G., Benzinger, T. L., . . . Brickman, A. M. (2016). White matter hyperintensities are a core feature of Alzheimer's disease: Evidence from the dominantly inherited Alzheimer network. *Ann Neurol, 79*(6), 929-939. doi:10.1002/ana.24647

Lepage, M., Habib, R., & Tulving, E. (1998). Hippocampal PET activations of memory encoding and retrieval: the HIPER model. *Hippocampus, 8*(4), 313-322. doi:10.1002/(SICI)1098-1063(1998)8:4<313::AID-HIPO1>3.0.CO;2-I

Libon, D. J., Glosser, G., Malamut, B. L., Kaplan, E., Goldberg, E., Swenson, R., & Sands, L. P. (1994). Age, executive functions, and visuospatial functioning in healthy older adults *Neuropsychology, 8*(1), 38-43.

Madden, D. J., Bennett, I. J., & Song, A. W. (2009). Cerebral white matter integrity and cognitive aging: contributions from diffusion tensor imaging. *Neuropsychol Rev, 19*(4), 415-435. doi:10.1007/s11065-009-9113-2

Malec, J. F., Ivnik, R. J., & Hinkeldey, N. S. (1991). Visual spatial learning test. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*(1), 82-88.

Mandecka, M., Budziszewska, M., Barczak, A., Peplonska, B., Chodakowska-Zebrowska, M., Filipek-Gliszczynska, A., . . . Gabryelewicz, T. (2016). Association between Cerebrospinal Fluid Biomarkers for Alzheimer's Disease, APOE Genotypes and Auditory Verbal Learning Task in Subjective Cognitive Decline, Mild Cognitive Impairment, and Alzheimer's Disease. *J Alzheimers Dis, 54*(1), 157-168. doi:10.3233/jad-160176

Meyers, J., & Meyers, K. (1995). The Myers Scoring System for the Rey Osterrieth Complex Figure: Professional Manual. *Odessa, FL: Psychological Assessment Resources*.

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology, 43*(11), 2412-2414.

Murphy, E. A., Holland, D., Donohue, M., McEvoy, L. K., Hagler, D. J., Jr., Dale, A. M., & Brewer, J. B. (2010). Six-month atrophy in MTL structures is associated with subsequent memory decline in elderly controls. *Neuroimage, 53*(4), 1310-1317. doi:10.1016/j.neuroimage.2010.07.016

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc, 53*(4), 695-699. doi:JGS53221 [pii]

10.1111/j.1532-5415.2005.53221.x

Norman, M. A., Evans, J. D., Miller, W. S., & Heaton, R. K. (2000). Demographically corrected norms for the California Verbal Learning Test. *J Clin Exp Neuropsychol, 22*(1), 80-94. doi:10.1076/1380-3395(200002)22:1;1-8;FT080

Oosterhuis, H. E., van der Ark, L. A., & Sijtsma, K. (2016). Sample Size Requirements for Traditional and Regression-Based Norms. *Assessment, 23*(2), 191-202. doi:10.1177/1073191115580638

Osterrieth, P. A. (1944). Le test de copie d'une figure complexe. *Archives de Psychologie, 30*, 206-356.

Palmqvist, S., Zetterberg, H., Blennow, K., Vestberg, S., Andreasson, U., Brooks, D. J., . . . Hansson, O. (2014). Accuracy of brain amyloid detection in clinical practice using cerebrospinal fluid beta-amyloid 42: a cross-validation study against amyloid positron emission tomography. *JAMA Neurol, 71*(10), 1282-1289. doi:10.1001/jamaneurol.2014.1358

Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Jr., Chance, J. M., & Filos, S. (1982). Measurement of functional activities in older adults in the community. *Journal of Gerontology, 37*(3), 323-329.

Prins, N. D., van Dijk, E. J., den Heijer, T., Vermeer, S. E., Jolles, J., Koudstaal, P. J., . . . Breteler, M. M. (2005). Cerebral small-vessel disease and decline in information processing speed, executive function and memory. *Brain, 128*(Pt 9), 2034-2041. doi:10.1093/brain/awh553

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., & Boesiger, P. (1999). SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine, 42*(5), 952-962.

Ranganath, C., Cohen, M. X., Dam, C., & D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *J Neurosci, 24*(16), 3917-3925. doi:10.1523/JNEUROSCI.5053-03.2004

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique.(Les problems.). *Archives de psychologie*.

Rey, A. (1968). *Epreuves Mnesiques et d'apprentissage*. Paris: Delachaux and Niestle.

Samgard, K., Zetterberg, H., Blennow, K., Hansson, O., Minthon, L., & Londos, E. (2010). Cerebrospinal fluid total tau as a marker of Alzheimer's disease intensity. *Int J Geriatr Psychiatry, 25*(4), 403-410. doi:10.1002/gps.2353

Schacter, D. L., Cooper, L. A., & Delaney, S. M. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *Journal of Experimental Psychology General, 119*(1), 5-24.

Schmidt, P., Gaser, C., Arsic, M., Buck, D., Forschler, A., Berthele, A., . . . Muhlau, M. (2012). An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *Neuroimage, 59*(4), 3774-3783. doi:10.1016/j.neuroimage.2011.11.032

Shirk, S. D., Mitchell, M. B., Shaughnessy, L. W., Sherman, J. C., Locascio, J. J., Weintraub, S., & Atri, A. (2011). A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. *Alzheimers Res Ther, 3*(6), 32. doi:10.1186/alzrt94

Sivan, A. B. (1992). *Benton visual retention test*: Psychological Corporation San Antonio, TX.

Thompson, P. M., Hayashi, K. M., De Zubicaray, G. I., Janke, A. L., Rose, S. E., Semple, J., . . . Toga, A. W. (2004). Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage, 22*(4), 1754-1766. doi:10.1016/j.neuroimage.2004.03.040

Tierney, M., Nores, A., Snow, W., Fisher, R., Zorzitto, M., & Reid, D. (1994). Use of the Rey Auditory Verbal Learning Test in differentiating normal aging from

Alzheimer's and Parkinson's dementia. *Psychological Assessment, 6*(2), 129-134. doi:10.1037/1040-3590.6.2.129

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale- 4th Edition*. San Antonio TX Pearson.

Wechsler, D. (2009). *Wechsler Memory Scale-Fourth Edition* (Fourth Edition ed.). San Antonio, TX: Pearson.

Wilkinson, G. S. (1993). *Wide Range Achievement Test-3 (WRAT-3) Administration Manual*. San Antonio, TX: The Psychological Corporation.

Wolk, D. A., & Dickerson, B. C. (2011). Fractionating verbal episodic memory in Alzheimer's disease. *Neuroimage, 54*(2), 1530-1539. doi:10.1016/j.neuroimage.2010.09.005

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research, 17*(1), 37-49.

Zetterberg, H., Skillback, T., Mattsson, N., Trojanowski, J. Q., Portelius, E., Shaw, L. M., . . . Blennow, K. (2016). Association of Cerebrospinal Fluid Neurofilament Light Concentration With Alzheimer Disease Progression. *JAMA Neurol, 73*(1), 60-67. doi:10.1001/jamaneurol.2015.3037

Table 1. Neuropsychological Protocol

Domain	Test	Description	Range
Global Cognition	Montreal Cognitive Assessment (Nasreddine et al., 2005)	Measures global cognitive status	0-30
Episodic Learning & Memory	Biber Figure Learning Test (BFLT; Glosser et al., 2002; Glosser et al., 1989)		
	BFLT Total Learning (Trials 1-5)	Assesses learning for a set of 15 geometric designs across 5 trials	0-45
	BFLT Distractor Trial (Trial B)	Assesses interference of learning a similar, novel set of 15 geometric designs	0-45
	BFLT Short Delay Free Recall	Assesses immediate free recall for a set of 15 geometric designs following 5 learning trials and presentation of a 15-item distractor trial (without re-exposure to the original 15 test items)	0-45
	BFLT Long Delay Free Recall	Assesses delayed recall for a set of 15 geometric designs following a 20-minute filled delay	0-45
	BFLT Long Delay Recognition Total Correct	Total number of correctly recognized geometric designs following a 20-minute filled delay	0-15
	BFLT Long Delay Recognition Distractor Trial False Alarms	Number of Distractor Trial geometric designs endorsed following a 20-minute filled delay	0-7
	BFLT Long Delay Recognition Novel False Alarms	Number of novel geometric designs endorsed following a 20-minute filled delay	0-23
	BFLT Long Delay Recognition Total False Alarms	Total number of non-target geometric designs recognized following a 20-minute filled delay	0-30
	BFLT Long Delay Recognition Total Discrimination	Assesses ability to discriminate the list of 15 geometric designs from Distractor and non-target designs after a 20-minute filled delay using the following formula: ((Recognition Total Correct + 0.5)/16))- ((Total False Alarms + 0.5)/31))	-30-15

	BFLT Repetitions	Total number of geometric designs repeated during Total Learning (Trials 1-5), Distractor Trial, Short Delay Recall, Long Delay Recall	n/a
	BFLT Intrusions	Total number of non-target geometric designs endorsed during Total Learning (Trials 1-5), Distractor Trial, Short Delay Recall, Long Delay Recall	n/a
	California Verbal Learning Test-II (CVLT-II; Delis et al., 2000)		
	CVLT-II Total Learning (Trials 1-5)	Assesses learning for a list of 16 words across 5 trials	0-16
	CVLT-II Distractor Trial	Assesses interference of learning a similar, novel list of 16 words	0-16
	CVLT-II Short Delay Free Recall	Assesses short delay free recall for a list of 16 words following 5 learning trials and presentation of a 16-item distractor trial (without re-exposure to the original 16 test items)	0-16
	CVLT-II Long Delay Free Recall	Assesses delayed free recall for a list of 16 words after a 20-minute filled delay	0-16
	CVLT-II Recognition Total Discrimination	Assesses ability to recognize the list of 16 words from related and non-related non-target words after a 20-minute filled delay	-4-4
	Benton Visual Retention Test 5th Edition (BVRT; Benton, 1974) Administration A Form C*	Assesses immediate visual memory for 10 designs presented for 10 seconds before reproducing them	0-10
Visuospatial Skills	Wechsler Adult Intelligence Scale-4th Edition Block Design (WAIS-IV; Wechsler, 2008)*	Assesses the ability to visuospatial organization and construction	0-66
	Hooper Visual Organization Test (HVOT; Hooper, 1983)	Measures proficiency of object recognition	0-30
Language	Boston Naming Test-30 Item (even items; Kaplan, Goodglass, & Weintraub, 1983)	Assesses confrontation naming and lexical retrieval abilities	0-30

	Category Fluency (Animal Naming; Goodglass & Kaplan, 1983)	Measures rapid word generation in 60 seconds based on a specified category	n/a
Information Processing	DKEFS Trail Making Test Number Sequencing (Delis et al., 2001)†	Measures visual-scanning and attention in a number sequencing task	0-150
	WAIS-IV Coding (Wechsler, 2008)	Speeded measure assessing psychomotor, attention and processing speed	0-93
Executive Functioning	DKEFS Tower Test (Delis et al., 2001)	Measures planning and problem solving abilities	0-30
	DKEFS Color-Word Interference Test Inhibition (Delis et al., 2001)†	Measures inhibition involving suppression of an automatic response in favor of a novel response	0-90
	DKEFS Trail Making Test Number-Letter Switching (Delis et al., 2001)†	Measures sequencing and mental flexibility in a number and letter set-shifting task	0-240
	Letter Fluency (Controlled Oral Word Association; Benton, Hamsher, & Sivan, 1994)	Measures rapid word generation based on a specified letter across three trials (F, A, and S), each lasting 60 seconds	n/a
Mood	Geriatric Depression Scale (GDS)-30 Item (Yesavage et al., 1983)	Assesses symptoms of depressed mood	0-30
Estimated Pre-Morbid Intelligence	Wide Range Achievement Test-3rd Edition (WRAT-3) Reading Subtest (Wilkinson, 1993)*	Measures reading for words with irregular sound-to-spelling correspondence	0-57

Note. *Measure administered at the eligibility visit. All other tests administered at enrollment visit. †Speeded test where time to completion is the outcome and higher score denotes worse performance. DKEFS= Delis Kaplan Executive Function System.

Table 2. Participant Characteristics

	NC	Early MCI	MCI	Combined Sample	p-value
Sample Size	174	27	131	332	--
Demographic Characteristics					
Age, years	72±7	73±6	73±8	73±7	0.66
Sex, % female	41	26	44	41	0.24
Race, % White/Non-Hispanic	87	85	85	86	0.88
Education, years	16±3	16±3	15±3	16±3	<0.001*§
Biber Figure Learning Test Performances					
Trial 1	14.4±5.4	11.1±5.2	8.6±5.4	11.8±6.1	<0.001*‡§¶
Trial 2	24.4±6.8	18.4±7.1	14.2±6.9	19.9±8.4	<0.001*‡§¶
Trial 3	29.4±7.3	23.9±6.9	17.9±7.7	24.4±9.2	<0.001*‡§¶
Trial 4	32.5±7.3	27.2±6.2	19.8±8.3	27.1±9.7	<0.001*‡§¶
Trial 5	34.5±7.1	29.4±5.9	21.8±9.4	29.1±10.0	<0.001*‡§¶
Total Learning (Trials 1-5)	135±31	110±28	82±35	112±41	<0.001*‡§¶
Distractor Trial (Trial B)	11.7±5.8	8.9±4.8	7.3±4.8	9.7±5.7	<0.001*§
Short Delay Free Recall	31.2±8.1	26.3±6.2	17.6±9.8	25.4±10.8	<0.001*‡§¶
Long Delay Free Recall	32.5±7.6	28.0±6.6	19.1±9.9	26.8±10.6	<0.001*‡§¶
Long Delay Recognition- Hits	14.0±1.5	13.2±1.6	12.5±2.4	13.4±2.0	<0.001*‡§
Long Delay Recognition- Related False Alarms	0.9±1.3	1.2±1.3	2.3±1.6	1.5±1.6	<0.001*§¶
Long Delay Recognition- Unrelated False Alarms	1.2±1.9	2.5±2.7	4.5±3.7	2.6±3.2	<0.001*‡§¶
Long Delay Recognition- Total False Alarms	2.1±2.8	3.7±3.7	6.8±4.9	4.1±4.4	<0.001*§¶
Long Delay Recognition- Discrimination	0.8±0.2	0.7±0.2	0.6±0.2	0.7±0.2	<0.001*‡§¶
Total Repetitions	0.2±0.5	0.2±0.4	0.3±0.8	0.2±0.6	0.88
Total Extraneous Responses	0.6±1.5	0.3±0.5	0.7±3.2	0.6±2.3	0.18
Brain MRI Measures					
Right Hippocampal Volume, mm^3	3875±437	3741±343	3613±468	3759±459	<0.001*§

Left Hippocampal Volume, mm^3	3537±425	3348±323	3243±467	3404±456	<0.001*§
Total Hippocampal Volume, mm^3	7412±840	7089±629	6856±878	7163±879	<0.001*§
Right Inferior Lateral Ventricle Volume, mm^3	807±439	964±652	1147±747	956±617	<0.001*§
Left Inferior Lateral Ventricle Volume, mm^3	869±471	1011±615	1254±836	1034±675	<0.001*§
Total Inferior Lateral Ventricle Volume, mm^3	1676±879	1974±1232	2401±1493	1990±1235	<0.001*§
Intracranial Volume, cm^3	1377±142	1406±106	1380±150	1381±142	0.47
Log White Matter Hyperintensity Volume	1.8±1.3	2.0±0.9	2.3±1.2	2.0±1.3	<0.001*§
Cerebrospinal Fluid Markers					
Amyloid-β_{42}, pg/mL	757±230	817±282	620±234	713±246	0.001*§
Total-tau, pg/mL	373±175	429±125	505±289	427±228	0.011*§
Phosphorylated-tau, pg/mL	56±22	63±17	68±31	61±26	0.049*
Neurofilament Light, pg/mL	936±453	1088±465	1253±718	1069±583	0.003*§

Note: Data presented as mean±standard deviation or frequencies; NC=cognitively normal control; MCI=mild cognitive impairment *p<0.05; ‡NC>early MCI; §NC>MCI; ¶early MCI>MCI

Table 3. Neuropsychological Performances in Cognitively Normal Participants

	n=174
Montreal Cognitive Assessment	27.0±2.2
CVLT-II Total Learning (Trials 1-5)	46.9±9.4
CVLT-II Distractor Trial	4.7±1.7
CVLT-II Short Delay Free Recall	10.0±3.3
CVLT-II Long Delay Free Recall	10.5±3.3
CVLT-II Recognition Total Discrimination	3.0±0.7
Benton Visual Retention Test-V Total Score	6.8±1.5
Wechsler Adult Intelligence Scale-IV Block Design	34.7±10.4
Hooper Visual Organization Test	25.3±2.5
Boston Naming Test 30-item (even items)	27.9±2.0
Category Fluency (Animals)	21.0±4.9
DKEFS Trail Making Test Number Sequencing*	36±13
Wechsler Adult Intelligence Scale-IV Coding	57±12
DKEFS Tower Test	16.1±4.3
DKEFS Color-Word Interference Test Inhibition*	60±14
Letter Fluency (FAS)	42.9±11.4
DKEFS Trail Making Test Number-Letter Switching*	87±34
Geriatric Depression Scale 30-item	2.4±2.8
Wide Range Assessment Test-3 Reading Subtest	51.3±4.3

Note: CVLT-II=California Verbal Learning Test; DKEFS=Delis Kaplan Executive Functioning System; *Higher score reflects worse performance

Table 4. Biber Figure Learning Test & Markers of Amyloid, Neurodegeneration, & White Matter Disease

	BFLT Total Learning		BFLT Short Delay Free Recall		BFLT Long Delay Free Recall		BFLT Discrimination	
	B	**p-value**	**B**	**p-value**	**B**	**p-value**	**B**	**p-value**
CSF $A\beta_{42}$	0.64	0.30	3.00	0.20	3.69	0.12	124.35	0.28
CSF T-tau	-1.44	0.01*	-6.70	0.002*	-7.84	<0.001*	-450.13	<0.001*
CSF P-tau	-0.15	0.02*	-0.70	0.005*	-0.80	0.001*	-43.31	<0.001*
CSF NFL	-3.03	0.03*	-8.00	0.14	-6.85	0.22	-307.32	0.25
Right Hippocampus Volume	1.07	0.10	5.39	0.03*	7.49	0.002*	152.90	0.18
Left Hippocampus Volume	2.25	<0.001*	9.36	<0.001*	11.40	<0.001*	377.99	0.001*
Total Hippocampal Volume	3.32	0.008*	14.75	0.001*	18.88	<0.001*	530.90	0.01*
Right Inferior Lateral Ventricle Volume	-1.41	0.14	-6.37	0.07	-7.19	0.05*	-497.16	0.003*
Left Inferior Lateral Ventricle Volume	-2.36	0.02*	-8.36	0.03*	-11.37	0.003*	-596.07	<0.001*
Total Inferior Lateral Ventricle Volume	-3.77	0.04*	-14.73	0.03*	-18.55	0.008*	-1093.23	<0.001*
WMH Volume	-0.002	0.26	-0.01	0.46	-0.01	0.50	-0.54	0.12

Note: BFLT=Biber Figure Learning Test; CSF= cerebrospinal fluid; Aβ=amyloid beta; T-tau=total tau; p-tau=phosphorylated tau; NFL=neurofilament light; WMH=white matter hyperintensity (log-transformed); *significance threshold at p-value<0.05.

Table 5. Biber Figure Learning Test Correlations with Other Neuropsychological Measures

	BFLT Total Learning		BFLT Short Delay Free Recall		BFLT Long Delay Free Recall		BFLT Discrimination	
	r	p-value	r	p-value	r	p-value	r	p-value
Montreal Cognitive Assessment	0.49	<0.001*	0.40	<0.001*	0.41	<0.001*	0.42	<0.001*
CVLT-II Total Learning (Trials 1-5)	0.54	<0.001*	0.49	<0.001*	0.51	<0.001*	0.43	<0.001*
CVLT-II Distractor Trial	0.26	<0.001*	0.30	<0.001*	0.22	0.004*	0.13	0.09
CVLT-II Short Delay Free Recall	0.51	<0.001*	0.50	<0.001*	0.50	<0.001*	0.44	<0.001*
CVLT-II Long Delay Free Recall	0.53	<0.001*	0.54	<0.001*	0.52	<0.001*	0.45	<0.001*
CVLT-II Recognition Discrimination	0.36	<0.001*	0.37	<0.001*	0.36	<0.001*	0.34	<0.001*
BVRT Total Score	0.41	<0.001*	0.44	<0.001*	0.44	<0.001*	0.24	0.002*
WAIS-IV Block Design	0.33	<0.001*	0.32	<0.001*	0.34	<0.001*	0.36	<0.001*
Hooper Visual Organization Test	0.38	<0.001*	0.41	<0.001*	0.31	<0.001*	0.29	<0.001*
Boston Naming Test 30-Item	0.38	<0.001*	0.38	<0.001*	0.30	<0.001*	0.36	<0.001*
Category Fluency (Animals)	0.32	<0.001*	0.37	<0.001*	0.31	<0.001*	0.28	<0.001*
DKEFS Number Sequencing†	-0.27	<0.001*	-0.33	<0.001*	-0.31	<0.001*	-0.24	0.001*
WAIS-IV Coding	0.41	<0.001*	0.42	<0.001*	0.40	<0.001*	0.26	<0.001*
DKEFS Tower Test	0.25	<0.001*	0.23	0.002*	0.23	0.002*	0.18	0.02*
DKEFS Color-Word Inhibition†	-0.27	<0.001*	-0.22	0.004*	-0.18	0.02*	-0.13	0.09
DKEFS Number-Letter Switching†	-0.28	<0.001*	-0.31	<0.001*	-0.30	<0.001*	-0.26	<0.001*
Letter Fluency (FAS)	0.41	<0.001*	0.37	<0.001*	0.28	<0.001*	0.36	<0.001*
Geriatric Depression Scale 30-Item	0.07	0.37	0.06	0.47	0.04	0.58	-0.04	0.59

Note: CVLT-II=California Verbal Learning Test-2nd Edition; BVRT=Benton Visual Retention Test 5th Edition; WAIS-IV=Wechsler Adult Intelligence Scale-4th Edition; DKEFS=Delis Kaplan Executive Functioning System, *significance threshold at p<0.01; †higher scores reflect worse performance

Table 6. Mean and Regression Coefficients for Normative Data Calculation

	Mean±SD	Intercept	β (age)	β (sex)	β (race/ ethnicity)	β (education)	RMSE
BFLT Total Learning (Trials 1-5)	135.0±31.0	223.7	-1.79***	12.78**	-10.30	2.26*	27.25
BFLT Distractor Trial	11.7±5.8	21.6	-0.21***	2.09*	1.68	0.23	5.49
BFLT Short Delay Free Recall	31.2±8.1	59.9	-0.46***	1.26	-4.75**	0.30	7.26
BFLT Long Delay Free Recall	32.5±7.6	54.4	-0.40***	1.11	-3.59*	0.45	6.96
BFLT Recognition Discrimination	0.82±0.15	0.9	-0.004**	0.03	-0.08*	0.01**	0.14

Note: *p<0.05, **p<0.01, ***p<0.001; RMSE=root mean square error