Associations Between Episodic Memory and Hippocampal Volume in Late Adulthood

by

Sarah Lillian Aghjayan

Bachelor of Arts, University of Vermont, 2014

Master of Science, University of Pittsburgh, 2019

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Sarah Lillian Aghjayan

Defended on

April 19, 2023

and approved by

Thomas W. Kamarck, Professor, Department of Psychology, University of Pittsburgh

Anna L. Marsland, Professor, Department of Psychology, University of Pittsburgh

Chaeryon Kang, Assistant Professor, Department of Biostatistics, University of Pittsburgh

Michelle W. Voss, Associate Professor, Department of Psychology, University of Iowa

Dissertation Chair: Kirk I. Erickson, Professor, Department of Psychology, University of Pittsburgh

Associations Between Episodic Memory and Hippocampal Volume in Late Adulthood

Sarah Lillian Aghjayan, Ph.D.

University of Pittsburgh, 2024

There are various ways of conceptualizing and assessing episodic memory (EM), but different EM tasks are only moderately correlated with each other, suggesting that EM might not be a unitary construct. Further, various EM tasks exhibit disproportional task demands on the hippocampus and differentially reflect hippocampal volume (HV) degeneration – one of the strongest predictors of Alzheimer's disease. Therefore, it is unclear if variation in EM performance is a meaningful indicator of risk for developing Alzheimer's disease. This study established a structural equation model to examine the covariance structure and distinctiveness of EM tasks and assessed whether these relate differently to HV. This study examined 648 older adults (M=69.88). EM was assessed using seven of the most commonly used tasks in neuropsychological testing settings. Automated Segmentation of Hippocampal Subfields was used to segment the hippocampus. A confirmatory factor analysis was used with residual covariances included and loadings freely estimated. Hierarchical regression models were used to test the associations between the observed factors of EM and HV. A model with three first-order subfactors (verbal immediate recall, verbal delayed recall, and visuospatial) derived from a second-order general EM domain factor had the best model fit. All three subfactors and the general EM domain factor significantly explained a similar amount of variance in total, left, and right HV. In addition, all subfactors were significantly associated with CA1, entorhinal cortex, and subiculum volume, only the verbal immediate recall and verbal delayed recall subfactors were significantly associated with CA3 volume, and none of the three subfactors were significantly associated with CA2 or dentate

gyrus volume. These results suggest that traditional EM tasks are measuring the same overarching construct, but different task conditions are tapping into different complex processes associated with EM. Further, performance across the observed factors only accounted for a small portion of the variance in HV, suggesting that HV might not be a strong marker of EM ability before clinically observable cognitive deficits are present. Lastly, findings from this study suggest that different hippocampal subfields are not uniformly involved in managing and supporting EM, and they may be preferentially important for various processes.

Table of Contents

List of Tables

List of Figures

1.0 Background and Significance

When you pass a car marked with the "student driver" sticker, does it prompt you to remember a time when *you* were first learning to drive? You might remember feeling excited sitting behind the wheel for the first time as your driving instructor described the controls of the car. Your ability to recall this specific, personal event – the people, places, and feelings you experienced at a certain moment in time and space – is known as episodic memory (Madan, 2020). As we age, episodic memory is one of the earliest cognitive domains to decline (Tulving, 2002). Further, a deficit in episodic memory is one of the first clinically observable symptoms among patients with Alzheimer's disease (Hodges et al., 2000). Deficits in episodic memory are associated with increases in social isolation (DiNapoli et al., 2014) and difficulties performing activities of daily living (Tomaszewski Farias et al., 2009), such as grocery shopping. Given that an estimated 6.2 million older adults in the United States are currently living with Alzheimer's disease (Rajan et al., 2021), there is a critical need to better understand which episodic memory measures are most reliable and predictive of future decline.

In research and clinical practice, there are various ways of conceptualizing and assessing episodic memory. Tasks that are classically used to assess episodic memory can include different materials (i.e., verbal or visual) and assess distinct processes (i.e., learning, immediate recall, or delayed recall). Additionally, there is a range of paradigms that attempt to assess various aspects of episodic memory function, such as memory capacity for semantically unrelated items (word list), recollection of logically linked ideas (story learning), or visuospatial memory (figure learning). These many different attributes of episodic memory tasks make it unsurprising that task performance is rarely/never independent from other cognitive processes. For example, sustained and selective attention (Aly & Turk-Browne, 2017) and executive functions, such as working

memory demands (Hill et al., 2012), influence the ability to successfully encode and retrieve information. In support of this, word list tasks are more strongly associated with executive dysfunction than story learning tasks, as executive functioning greatly affects the number of words learned across trials (Tremont et al., 2000). Thus, variation in performance on tasks that are commonly used for measuring episodic memory might not be entirely attributable to memory ability but rather to individual variation in other cognitive (i.e., non-episodic) processes.

In short, many widely used instruments designed for measuring "episodic memory" may not be uniformly or equivalently assessing episodic memory and are not independent from the contributions of other cognitive abilities. Consistent with this argument, there is a significant overlap in scores on tasks commonly used to assess episodic memory across the Alzheimer's disease spectrum (Bäckman et al., 2005). More specifically, the distribution of baseline scores on episodic memory tasks for individuals with normal cognition and individuals with subtle decline who later received an Alzheimer's disease diagnosis overlapped by 42.1% in a meta-analysis, suggesting that nearly half of the individuals with normal cognition had scores in the impaired range and vice versa (Bäckman et al., 2005). Given this heterogeneity, it is unclear if variation in performance on episodic memory measures is primarily due to variation in episodic memory ability or even whether it is a meaningful indicator of risk for developing Alzheimer's disease. A better understanding of which episodic memory measures are predictive of risk for Alzheimer's disease could have wide-reaching implications for clinical neuropsychologists, neurologists, and researchers, and it might allow for earlier detection of subtle changes in episodic memory. It might also allow clinicians to improve treatment recommendations to slow brain deterioration associated with cognitive decline and allow for more precise neuropsychological protocols for identifying individuals for pharmaceutical interventions and clinical trials.

1.1 Episodic Memory

Episodic memory performance is mediated by several complex processes, including encoding, consolidation, storage, and retrieval. Encoding episodic information involves the representation of different kinds of information (e.g., perceptual, semantic) linked together across a short period of time in space (Mayes & Roberts, 2001). Attention plays a critical role during encoding, as only components that are focused on will get into memory (Mayes & Roberts, 2001). It has been argued that meaningful encoding of episodic information, which will later enhance recall, depends on the successful processing of information through semantic memory (Tulving, 1995). A fraction of the information that is encoded undergoes an active consolidation process over time that fixes the memories into long-term storage, with a cascade of cellular and molecular processes occurring immediately after learning (Yonelinas et al., 2019). Standard systems consolidation theory argues that episodic memories will be forgotten unless they are consolidated gradually during offline periods (e.g., sleep) to become more fully represented in the neocortex (Squire et al., 2015). It is widely held that episodic memories are stored in the brain where they were originally represented, such that the same neural array will be activated when retrieving the information either immediately, hours, days, or years after the experience (Gaffan & Hornak, 1997), yet the way in which cellular analogs of memory translate into experienced memories remains a mystery. Lastly, retrieving episodic memories is typically an intentional and effortful directed search for a target memory and is accompanied by the feeling of familiarity (Mayes & Roberts, 2001). Whereas recall involves an organized search process to retrieve appropriate cues, recognition depends on both how familiar an item feels and whether anything else can be retrieved about the item or episode in which it appeared (Mayes & Roberts, 2001).

Examining the covariance structure between different episodic memory tasks and processes could provide insight as to whether the tasks are measuring the same constructs and the extent to which some of them could be used interchangeably. A study by Sudo and colleagues (2019) found that different materials and designs of episodic memory tasks are only moderately correlated with each other. More specifically, the correlation coefficient between performance on a word list task, a story learning task, and a figure learning task ranged from $r = 0.44$ to $r = 0.56$. However, this study consisted of a small sample size (N=27) and included individuals with normal cognition, mild cognitive impairment (the transitional phase between normal cognitive functioning and dementia; Petersen et al., 2018), and dementia due to Alzheimer's disease. Nevertheless, these results suggest that traditional measures that are commonly and jointly referred to as 'episodic memory' tasks may be measuring several subtypes of episodic memory rather than a single, general, overarching construct.

Several studies have conducted factor analyses to more precisely characterize the construct validity of episodic memory tasks. For example, when examining 6 tests across different cognitive functions (e.g., confrontation naming, verbal fluency, visual attention and task switching, visuomotor ability, verbal memory, and visuospatial memory), a verbal list learning memory test (Hopkins Verbal Learning Test-Revised; HVLT-R) loaded onto a single factor with only the visuospatial memory task (Brief Visuospatial Memory Test-Revised; BVMT-R) (Shapiro et al., 1999). However, in another factor analysis that examined similar measures, there was a verbal memory factor that encompassed multiple scores from the HVLT-R that was separate from the visual memory factor that encompassed multiple BVMT-R scores (Benedict et al., 1996). While these results together suggest that both of these verbal and visual memory tests are largely tapping into different cognitive abilities from other tests, it is unclear the extent that these traditional

episodic memory tasks are measuring a similar construct of episodic memory. A greater focus on memory measures could provide clarity as to whether there are separate subfactors of episodic memory that some instruments are measuring to a greater extent. A previous study focused on the general domain of memory, including tests of episodic memory and working memory, and found that the factors that emerged were closely linked to the tasks, such that one factor was tied to word list tasks, one to paired-associates tasks, and one to working memory tasks (Underwood et al., 1978). However, no study to date has conducted a factor analysis using only tasks that have been traditionally assumed to assess the unitary construct of episodic memory.

Since episodic memory tasks might be measuring distinct subfactors, a closer examination of cognitive performance among individuals without dementia who later receive an Alzheimer's disease diagnosis could provide insight as to which measures, and which subfactors, are most predictive of future decline. A meta-analysis found greater deficits across delayed performance than immediate performance on episodic memory tasks among individuals with subtle cognitive decline who later received an Alzheimer's disease diagnosis compared to older adults with normal cognition (Bäckman et al., 2005). These results are consistent with the theoretical view that transferring information from temporary storage to a more permanent memory representation is a cardinal feature of episodic memory (Squire, 1986). Bäckman and colleagues (2005) also found larger effect sizes for recall scores compared to recognition scores. These greater deficits in recall performance among individuals who later received an Alzheimer's disease diagnosis may be because retrieval is typically more demanding than recognition, as recognition can rely on general feelings of familiarity (Jacoby et al., 1993). Lastly, Bäckman and colleagues (2005) also found larger effect sizes for verbal tasks compared to visuospatial tasks. These results are in line with previous research that suggests retention of material presented verbally is worse than visual stimuli

(Bigelow & Poremba, 2014). Overall, these results suggest that delayed memory for semantically unrelated words presented verbally may be more cognitively demanding, and decreased performance may be a lead indicator of risk for future decline.

1.2 Episodic Memory and Hippocampal Volume

Episodic memory is supported by a distributed network of cortical and subcortical brain structures but requires the involvement of the hippocampus, which is located in the medial temporal lobe (Madan, 2020). During encoding, representations of episodic memories are comprised of complex neural activity patterns within the medial temporal lobe (Mayes, 1988), the frontal neocortex to control executive functions (Kapur et al., 1994), and various parts of the posterior neocortex depending on the nature of the experience (i.e., the precuneus is activated when encoding complex scene pictures) (Mayes & Montaldi, 1999). Initially, during consolidation, new protein synthesis and synapse changes occur in the medial temporal lobe and hippocampus, but over time the changes gradually occur primarily in the posterior neocortex so that the hippocampus is less necessary for retrieval due to the direct links that have developed to different neocortical regions (Squire & Alvarez, 1995). During retrieval, the search and monitoring process is dependent on frontal neocortical regions (Mayes, 1988), recollection is primarily dependent on the hippocampus (Aggleton & Brown, 1999), and familiarity is dependent on the perirhinal cortex, dorsomedial thalamus, and various parts of the frontal neocortex (O'Reilly & Rudy, 2001). Thus, the hippocampus is involved to some extent across all the complex processes of episodic memory.

The hippocampus is more susceptible to age-related deterioration than other brain regions among older adults with (Frisoni et al., 2010) and without (Raz et al., 2005) Alzheimer's disease. Both cross-sectional and longitudinal studies report reduced hippocampal volume in patients with

mild cognitive impairment and Alzheimer's disease compared to healthy controls (Shi et al., 2009). Further, recent evidence suggests that smaller hippocampal volume at baseline is one of the strongest predictors of a faster decline in episodic memory and conversion to Alzheimer's disease – even more so than some of the hallmark biomarkers of Alzheimer's disease pathology, such as tau and amyloid (Ottoy et al., 2019). However, it remains unclear which episodic memory measures are most sensitive to variation in hippocampal volume before cognitive deficits emerge.

Determining the association between episodic memory performance and hippocampal volume could help elucidate which measures best reflect neurodegeneration within the spectrum of Alzheimer's disease. The aforementioned study by Sudo and colleagues (2019) found that scores on a word list task explained 35-48% of the variance in hippocampal atrophy, whereas hippocampal atrophy was not significantly correlated with a story learning task or a figure learning task. Although the authors did not directly compare the scores to examine which explains significantly greater variance in hippocampal volume, the findings indicate that memory for semantically unrelated words may be more strongly correlated with hippocampal volume decline than other episodic memory tasks. These results, in combination with those found by Bäckman and colleagues (2005), suggest that delayed memory for semantically unrelated words presented verbally may be an important early predictor of risk for developing Alzheimer's disease. They also suggest that various episodic memory tasks exhibit disproportional task demands on the hippocampus and, therefore, should not be treated as equivalent measures. However, as mentioned previously, Sudo and colleagues (2019) included a small sample size comprised of individuals with cognitive decline and only examined three episodic memory tasks. Thus, they were not able to conclude whether there was some task-related feature that drove the correlation with hippocampal volume. Unfortunately, prior studies have not comprehensively evaluated episodic

memory, so there is little data available to determine the existence of possible subfactors. Further, if and why subfactors of episodic memory might differentially relate to hippocampal volume in late adulthood is poorly understood.

The hippocampus itself is not a cellularly homogenous brain structure and its various subfields might perform different computations during episodic memory tasks. That is, the hippocampus is composed of three cornu ammonis (CA) regions (CA1, CA2, and CA3) and the dentate gyrus, it receives input from the entorhinal cortex, and its output travels via the subiculum and fimbria/fornix (Langston et al., 2010). There are several theories about the different functions of each subregion. For example, the dentate gyrus might act primarily as a pattern separator (O'Reilly & Rudy, 2001), CA1 as a mismatch detector and to add temporal context to events (Lisman & Otmakhova, 2001; Rolls & Kesner, 2006), and CA3 for pattern completion (Nakazawa et al., 2002). Further, atrophy in the subfields is not homogenous across the spectrum of Alzheimer's disease. For example, Carlesimo and colleagues (2015) found that CA1 does not display atrophic changes, subiculum volume progressively declines from mild cognitive impairment to Alzheimer's disease, while CA2-3 and dentate gyrus volumes decline in those with mild cognitive impairment but remain relatively stable in the progression to Alzheimer's disease.

These results suggest that the different subregions may not be uniformly involved in managing and supporting episodic memory processes. In fact, numerous studies have examined the association between the volumes of the hippocampal subfields and performance on episodic memory tasks and found differences in the strength of the correlations. For example, a study of adults across the lifespan found that bilateral CA1-3 and dentate gyrus volumes were associated with immediate recall performance on a list learning task, while bilateral CA2-3 and dentate gyrus volumes were associated with delayed recall scores (Zheng et al., 2018). In contrast, Mueller and

colleagues (2011) found that only CA3 and dentate gyrus volumes were associated with immediate recall performance whereas CA1 volume was associated with delayed recall scores in a sample of individuals with normal cognition and mild cognitive impairment. These contradictory results may be due to differences in study samples and the use of a single task to assess episodic memory. Examining hippocampal subfield volumes in relation to subfactors of episodic memory performance could help further elucidate the regions involved in specific episodic memory functions. However, no study to date has examined hippocampal subfield volume in relation to factor structures across different episodic memory tasks.

1.3 The Current Study

This study addresses several limitations of prior studies by using a much larger sample size with a more comprehensive evaluation of episodic memory to examine the covariance structure and distinctiveness of tasks that have been traditionally used as measures of episodic memory and assess whether these relate differently to hippocampal volume. A larger sample size provides greater power to conduct a factor analysis and detect distinct subfactors of episodic memory, examine the nature of each component (i.e., material, design), and determine which measures are most sensitive to variation in hippocampal volume. This study is the first to address the following three aims:

Aim 1) Use a factor analytic approach to describe a set of distinct subfactors of episodic memory across seven of the most commonly used tasks. Based on theories of episodic memory processes and the study by Bäckman and colleagues (2005), it is predicted that there will be four subfactors: verbal immediate recall, verbal delayed recall, visuospatial,

and recognition (see Figure 1). This proposed factor structure is comprised of both process factors (e.g., recall, recognition) and content factors (e.g., verbal, visuospatial).

Aim 2) Explore whether the subfactors explain similar variance in hippocampal volume as compared to the general domain factor of episodic memory or whether a distinct subfactor explains greater variation in hippocampal volume. Based on theories of episodic memory processes and the aforementioned study by Sudo and colleagues (2019), it is predicted that the verbal delayed recall component will explain the most variance in hippocampal volume, above and beyond the general domain of episodic memory and the other subfactors.

Aim 3) Explore whether the episodic memory subfactors explain similar variance across all hippocampal subfield volumes. It is predicted that the subfields will vary in their involvement in episodic memory subfactors, but because of inconsistencies across results in prior literature, this aim will remain exploratory.

With the United States population of older adults expected to reach 88 million by 2050 (He et al., 2016), the economic and social burdens associated with Alzheimer's disease will continue to grow. Therefore, a more comprehensive understanding of which episodic memory measures are closely related to hippocampal volume is imperative to more accurately identify individuals at risk of developing Alzheimer's disease. The findings generated by this study of cognitively normal older adults will lay the groundwork for determining which tasks and scores across episodic memory measures are most correlated with a critical Alzheimer's disease biomarker before a decline in cognition is clinically detectible. Early detection allows for the implementation of treatment options and enrollment in clinical trials sooner in the disease course, which has the potential to improve patient quality of life, minimize healthcare expenditures, and reduce the burden placed on public health resources.

2.0 Methods

2.1 Participants

This study is a secondary analysis of baseline data from a sample of 648 cognitively normal older adults who enrolled in a multi-site 12-month aerobic exercise intervention called Investigating Gains in Neurocognition in an Intervention Trial of Exercise (IGNITE; PI: Kirk Erickson, NCT02875301, R01AG053952). Participants were enrolled across three sites (University of Pittsburgh, Northeastern University, and University of Kansas) prior to March 2022 on a rolling basis. Participants were considered eligible if they met all of the following criteria: 65-80 years of age; ambulatory without pain or the use of assisted walking devices; ability to speak and read English; medical clearance to exercise by a primary care physician; living in the community for the duration of the study (i.e., able to travel to the exercise facility three times per week); reliable means of transportation; no diagnosis of a neurological condition (e.g., Parkinson's disease, dementia, stroke, multiple sclerosis); eligible to undergo magnetic resonance imaging (i.e., no metal implants that are MR ineligible, not claustrophobic); physically inactive consisting of engagement in less than 20 minutes of moderate-intensity physical activity per day for 2 days or less per week; Telephone Interview of Cognitive Status score > 25 (Brandt, 1991); and cognitive adjudication decision of cognitively normal using the 2011 National Institute on Aging-Alzheimer's Association criteria (Albert et al., 2011; McKhann et al., 2011). This study was approved by the Institutional Review Board, and participants provided written informed consent (Erickson et al., 2019).

2.2 Measures

2.2.1 Montreal Cognitive Assessment

The Montreal Cognitive Assessment (MoCA) is a brief cognitive screening tool that has high test-retest reliability ($r = 0.92$), internal consistency ($\alpha = 0.83$), and sensitivity and specificity for detecting mild cognitive impairment (87-90%) (Nasreddine et al., 2005). This 30-point test assesses eight cognitive domains, including short-term memory recall, visuospatial, executive functioning, language, attention and working memory, and orientation to time and place. This study focused on the short-term memory scores (5 points). Participants learned five nouns over two learning trials and were asked to repeat each word. After approximately 5 minutes, participants were instructed to freely recall the words, losing one point for each word incorrectly recalled. This study focused on raw free recall scores.

2.2.2 Logical Memory

The Logical Memory subtest from the Wechsler Memory Scale (Wechsler, 1997) was administered as part of the Virginia Cognitive Aging Project (Salthouse et al., 1996) in this study. It is a standardized and reliable ($\alpha > 0.7$) instrument used to assess contextual episodic memory (Salthouse, 2014). Participants were orally presented with two narrative stories, one presented once (Story A) and the other presented twice (Story B). They were asked to freely recall each story immediately after presentation (25 points each; 75 points total) and after a 20-minute delay (25 points each; 50 points total). This study focused on raw total immediate and delayed recall scores.

2.2.3 Paired Associates

A modified version of the Paired Associates subtest from the Wechsler Memory Scale (Wechsler, 1997) was administered as part of the Virginia Cognitive Aging Project (Salthouse et al., 1996) in this study. It is a standardized and reliable ($\alpha > 0.7$) instrument used to assess episodic memory (Salthouse, 2014). Participants were orally presented with two separate word lists, each containing six different word pairs, at a rate of approximately one word per second, with a longer pause between words from different pairs. For example, List 1 consisted of the word pair bell-pencil, and List 2 consisted of the word pair bank-clown. Participants were immediately presented with the first word in each pair and instructed to freely recall the response word associated with each stimulus word. Participants were allotted approximately 5-10 seconds per list item to recall the word. Immediate recall performance was assessed as the mean number of words recalled across both lists (maximum 6). After a 20-minute delay, participants were again presented with the first word in each pair and instructed to freely recall the response word associated with each stimulus word. Delayed recall performance was assessed as the mean number of words recalled across both lists (maximum 6). This study focused on mean immediate and delayed recall raw scores.

2.2.4 Hopkins Verbal Learning Test-Revised

The Hopkins Verbal Learning Test (HVLT) (Brandt, 1991) was originally designed to measure verbal episodic memory and has been revised (HVLT-R) to include a delayed recall trial (Brandt & Benedict, 2001). It has been found to have moderate to high reliability ($r = 0.55 - 0.78$) (Benedict et al., 1998). The word list consists of 12 items, four from three semantic categories (animals, stones, shelter), and was read to participants at the rate of approximately one word every

two seconds. Participants were then instructed to freely recall the words. The same procedure was repeated for two more trials, resulting in a total recall score across all three learning trials (36 total points). After a 20-minute delay, participants were again instructed to freely recall the words (12 points). After the delayed recall trial, participants were read 24 words and asked to respond after each word whether it was on the list (target) or not (distractor) (24 points). Half of the distractor words were from the same semantic categories as the target words and half were from unrelated categories. A recognition discrimination index was calculated by subtracting the number of false positives from true positives. This study focused on total learning, delayed recall, and recognition discrimination index raw scores.

2.2.5 Picture Sequence Memory Test

The NIH Toolbox includes the Picture Sequence Memory Test (PSMT), a visual episodic memory task (Zelazo et al., 2013). It has been found to have good test-retest reliability (r = 0.84) (Dikmen et al., 2014). Using an iPad, colored pictures of objects and activities were presented one at a time in a fixed, sequential order. The content of each picture was named/labeled orally simultaneously using a recording, although a trained examiner was present throughout the entire test session. Once a picture was presented, it was reduced in size and moved to its proper position in the sequence. Presentation, description, and placement in its correct location took approximately five seconds for each picture. The next picture was then presented without a delay, which continued until all pictures were displayed and placed in their fixed position. Then the pictures were scrambled into a random spatial arrangement and the participants were instructed to move each picture to its precise location to replicate the correct sequence. There was an initial introductory exercise of moving objects around the screen and then a brief (4-item) practice

sequence to orient participants to the task. After the practice, two picture sequences of a certain theme were presented to the participants, with the second sequence including additional items to the first. Raw scores reflect the cumulative number of adjacent pairs of pictures remembered correctly over the three trials.

2.2.6 Brief Visuospatial Memory Test-Revised

The Brief Visuospatial Memory Test-Revised (BVMT-R) is a widely used measure of visual episodic memory, with excellent interform reliability ($p > .05$) (Benedict et al., 1996). Participants were shown an 8x11 inch page for 10 seconds containing six geometric visual designs in a 2x3 matrix. Immediately following the presentation, participants were asked to draw as many designs as possible in the correct location and as accurately as possible without time constraints. Each design was awarded one point for correct location and one point for drawing accuracy (12 points). The same procedure was repeated for two more trials, resulting in a total recall score across all three learning trials (36 total points). Following a 25-minute delay, participants were asked to reproduce the matrix once again from memory (12 points). After the delayed recall trial, participants were presented with 12 designs and asked to respond after each design whether it was part of the matrix (target) or not (foil). Half of the designs were targets and half were foils. A recognition discrimination index was calculated by subtracting the number of false alarms from hits. This study focused on total recall, delayed recall, and recognition discrimination index raw scores.

2.2.7 Cohen's Relational Memory Test

An important component of episodic memory is relational memory, or the memory of the relationships between elements, such as face-scene pairs or object-location arrays (Eichenbaum & Cohen, 2004; Ngo et al., 2018). The hippocampus is known to play a critical role in relational memory and hippocampal amnesic patients are impaired on tasks assessing relational memory – across both short and long delays (Yee et al., 2014). Relational memory was assessed using a spatial reconstruction task developed by Neal Cohen and colleagues (Monti et al., 2015). Similar tasks have been used in previous studies to assess hippocampal-dependent memory performance (Jeneson et al., 2010; Watson et al., 2013). Using a computer, participants studied the arrangement of five separate line drawings for a fixed time period (10 seconds), after which the objects disappeared. Following an approximate 2-second delay, the stimuli reappeared aligned at the top of the screen, and participants were asked to click and drag each stimulus into the position where they thought it was positioned during the study phase with no time constraints. There were two seconds between each trial. There were three practice trials and 15 trials. This study focused on four scores (misplacement, edge resizing, distortion, and swaps) that have been shown to be highly sensitive to the structural integrity of the hippocampus (Monti et al., 2015; Watson et al., 2013). Misplacement measures the distance in centimeters between each item's studied location and the location where it was placed during reconstruction. Edge resizing measures reconstructed changes in the length in centimeters and direction in radians of vectors between each pair of items. Distortion measures the frequency of categorical changes in shape (e.g., changing a line into a square). A swap occurs when participants place two objects in spatial locations that were previously occupied in the study phase but not for the specific object of the current trial, and the

final swap score is the number of swap errors divided by the number of possible pairwise relations in a trial. For all scores, a lower value indicated better performance.

2.2.8 Magnetic Resonance Imaging (MRI)

Magnetic resonance images were collected at the University of Pittsburgh and Northeastern University using a Siemens Prisma 3T scanner with a 64-channel head coil and at the University of Kansas Medical Center using a Siemens Skyra 3T scanner with a 32-channel head coil. Imaging protocols were designed to be exact matches across the two scanner types. Before enrolling participants in the study, the image sequences and image quality were standardized across sites. Phantom scans were run at each site on a regular basis to ensure the stability of the data quality and geometric accuracy of the MRI scanners. A human phantom (GAG) was also scanned annually at each site. Images from a high-resolution T1-weighted 3D MPRAGE (Magnetization Prepared Rapid Gradient Echo Imaging) sequence (0.8 x 0.8 x 0.8 mm voxels, 224 slices, 0.8 mm slice thickness, TR = 2400.0 ms, TE = 2.31 ms, flip angle = 8 degrees) and T2-weighted focal hippocampal sequence (0.4 x 0.4 x 2.0 mm voxels, 30 slices, 2.0 mm slice thickness, TR = 8830.0 ms, TE = 78 ms, flip angle = 150 degrees) were collected for hippocampal subfield segmentation. A semi-automated software package called Automated Segmentation of Hippocampal Subfields (ASHS) was used to segment the subfields of the hippocampus. ASHS uses multi-atlas segmentation and machine learning techniques to identify and label the subfields of the hippocampus and medial temporal lobe cortices, and it has been shown to have good consistency with manual segmentation (Dice similarity coefficient = 0.5 - 0.8) (Yushkevich et al., 2015). For each participant, ASHS was used to generate measures of the volume of the subfields that make up the hippocampus: CA1, CA2, CA3, dentate gyrus, entorhinal cortex, and subiculum. Total, left,

and right hippocampal volume was calculated by summing the volume of the CA1, CA2, CA3, dentate gyrus, and subiculum. Estimated total intracranial volume was determined with FreeSurfer by registering images to an average template using a full 12-parameter affine transformation (Buckner et al., 2004). This approach is considered a robust method that is equivalent to manual correction in aging and dementia research (Buckner et al., 2004).

2.3 Procedures

Participants completed two separate cognitive assessment sessions before enrollment into the study. The first cognitive assessment lasted approximately 2.5 hours and included the MoCA, HVLT-R, and BVMT-R. Participants returned for a second 2.5-hour cognitive assessment session, during which Logical Memory, Paired Associates, PSMT, and Cohen's Relational Memory were collected. After the second cognitive assessment session, participants completed a separate MRI session. All baseline visits were completed within 8 weeks. This study utilized baseline data for participants who were randomized into the intervention.

2.4 Analyses

Categorical variables (e.g., gender, site) were treated as categorical in all analyses. All raw scores on episodic memory tasks were converted to *z*-scores. A higher value indicated better performance, except for Cohen's Relational Memory Test scores where a lower value reflected better performance. Preliminary analyses were conducted to ensure that assumptions of normality, linearity, missing data, and outliers were not violated. Skewness and kurtosis were examined using the R package moments (D′Agostino, 1970). Missing data were handled using full information maximum likelihood, eliminating the need for case-wise deletion. Bartlett's Sphericity Test

assessed the probability that at least some of the variables were significantly correlated using the R package performance (Lüdecke et al., 2021). The Kaiser-Meyer-Olkin (KMO) statistic assessed the factorability of the data using the R package performance (Lüdecke et al., 2021). The KMO statistic, ranging from 0 to 1, predicted whether the data were likely to factor well given the correlations among the variables. Using Kaiser's guidelines (1974), a cutoff of KMO $\geq$ .60 was used.

To address Aim 1, a confirmatory factor analysis was used, guided by *a priori* theory about the subfactor structure proposed in Figure 1, using the R package lavaan (Rosseel, 2012). Specifically, a model with a general episodic memory domain factor and four subfactors was estimated: 1) a verbal immediate recall component derived from the Logical Memory total immediate recall score, Paired Associates immediate recall score, and HVLT-R total learning score; 2) a verbal delayed recall component derived from MoCA, Logical Memory, Paired Associates, and HVLT-R delayed recall scores; 3) a visuospatial component derived from PSMT cumulative number of pairs, BVMT-R total and delayed recall scores, and all Cohen's Relational Memory scores; and 4) a recognition component derived from HVLT-R and BVMT-R recognition discrimination index raw scores. Each criterion was unidimensional and loaded on only one factor. See Table 1 for a list of measures and conditions analyzed.

A confirmatory factor analysis was used with residual covariances included, which added task-specific covariance that allowed the scores within a task to correlate with one another, with loadings freely estimated (Eid et al., 2008). The latent scales of each episodic memory subfactor were identified by fixing the loading of a reference indicator to one. The variance of the general episodic memory domain factor was also fixed to one. The remaining pattern coefficients, factor variances, and factor covariances were freely estimated (Millsap, 2001). Likelihood-ratio tests

were used to assess the statistical significance of parameter estimates within a model, such that a model with parameters freely estimated was compared to a nested model with fixed parameters. The difference in the likelihood ratio chi-square (χ^2) between the two models indicated the difference in fit, with significance suggesting that the null hypothesis can be rejected and the models do not fit equally well (Kline, 2015). Indicators with factor loadings less than 0.55 were removed to improve model fit, as it has been suggested that a cut-off of 0.55 indicates good (30%) overlapping variance (Comrey & Lee, 1992). A model χ^2 *p*-value $\geq$ .05, a Bentler Comparative Fit Index (CFI) > .90, a Steiger-Lind Root Mean Squared Error of Approximation (RMSEA) <.08, and a Standardized Root Mean Residual (SRMR) < .08 indicated a good model fit (Hu & Bentler, 1999). A post hoc power analysis was conducted using the R package semPower (Moshagen & Erdfelder, 2016). The power achieved with a sample size of N = 648 was determined to detect misspecifications of a model corresponding to RMSEA = 0.08 and an alpha error = 0.05.

To examine Aim 2, the estimated values for the latent variables in the above model were extracted. An *a priori* power analysis was conducted using G*Power version 3.1.9.3 (Faul et al., 2009). Based on data from Sudo and colleagues (2019) (*N*=27), the minimum sample size was estimated using the smallest effect they reported (r = 0.29), a significance criterion of α = 0.05, power = 0.80, and a two-tailed correlation. Bivariate Pearson correlations were examined between demographic variables (i.e., age, sex, race, education, and site), all episodic memory composite scores, and hippocampal volume. Statistically significant associations resulted in the inclusion of that demographic variable as a covariate in subsequent analyses. Preliminary analyses were conducted to ensure that assumptions of normality, linearity, missing data, and outliers were not violated among the composite scores using the R package moments (D′Agostino, 1970).

Hierarchical regression models were used to test the associations between the subfactors of episodic memory and hippocampal volume using the R package relaimpo (Groemping, 2007). Specifically, each subfactor composite score was entered into a single hierarchical regression model for total, left, and right hippocampal volume separately. The general regression model is displayed below:

Step 1: Hippocampal Volume = $\beta_0 + \beta_{covariates}$

Step 2: A) Hippocampal Volume = $\beta_0 + \beta_{covariates} + \beta_{EpisodicMemorySubfactor1}$

B) Hippocampal Volume = $\beta_0 + \beta_{covariates} + \beta_{EpisodicMemorySubfactor2}$

C) Hippocampal Volume = $\beta_0 + \beta_{covariates} + \beta_{EpisodicMemorySubfactor3}$

D) Hippocampal Volume = $\beta_0 + \beta_{covariates} + \beta_{EpisodicMemorySubfactor4}$

Step 3: Hippocampal Volume = $\beta_0 + \beta_{covariates} + \beta_{EpisodicMemorySubfactor1} +$

$\beta_{EpisodicMemorySubfactor2} + \beta_{EpisodicMemorySubfactor3} + \beta_{EpisodicMemorySubfactor4}$

Comparing the variance explained by each subfactor in step two (i.e., Step 2A vs. Step 2B vs. Step 2C vs. Step 2D) determined the marginal effect of each subfactor. Further, comparing the variance explained by each subfactor in step two of the model determined whether any one subfactor explained variance in hippocampal volume above and beyond the others. Significance was set *a priori* as $p < .05$.

Separate hierarchical linear regression models were used to test the association between the general domain factor score of episodic memory and hippocampal volume. Specifically, the general domain factor score was entered into a single regression model for total, left, and right hippocampal volume separately. The general regression model is displayed below:

Step 1: Hippocampal Volume = $\beta_0 + \beta_{covariates}$

Step 2: Hippocampal Volume = $\beta_0 + \beta_{covariates} + \beta_{EpisodicMemoryGeneralDpmainFactor}$

Comparing the amount of variance explained by the general episodic memory domain factor in step two with the amount of variance explained by the subfactor that emerged in the hierarchical regression model above allowed us to determine whether an individual subfactor explained greater variance in hippocampal volume than the general domain of episodic memory. Significance was set *a priori* as $p < .05$.

To examine Aim 3, hierarchical regression models were used to test the associations between the subfactors of episodic memory and the following hippocampal subfield volumes: 1) the ammonic subfields (CA1-3), 2) dentate gyrus, 3) entorhinal cortex, and 4) subiculum. Similar to Aim 2, each observed subfactor composite score was entered into a single hierarchical regression model for each subfield separately. Comparing the variance explained by each subfactor in step two of the model determined whether any one subfactor explained variance in hippocampal subfield volume above and beyond the others. Significance was set *a priori* as $p < .05$.

3.0 Results

3.1 Characteristics of the Sample

Means and standard deviations of demographic variables and episodic memory measures are presented in Table 2. The 648 participants were on average 69.88 years old (±3.75) with 16.32 years of education (±2.21). Females made up 71.1% of the sample and 75.8% were White. Of the measures that had demographically normed scores (i.e., HVLT-R, BVMT-R, PSMT), mean performance was in the average range or within 1.5 standard deviations (ranging between $t = 50.13$ (±11.17) and $t = 53.10$ (±9.00)). One participant was missing HVLT-R recognition discrimination index data, which was imputed using full information maximum likelihood. There were two outliers (HVLT-R recognition discrimination index = 1, BVMT-R recognition discrimination index = 1). However, the values did not reflect measurement error and, thus, were kept in all subsequent analyses. The variables were not heavily skewed and followed the normal distribution to an acceptable extent (absolute skewness range 0.1 – 2.0 and absolute kurtosis range 2.2 – 7.0) (Kline, 2015).

3.2 Confirmatory Factor Analysis

Correlations between the episodic memory measures are presented in Table 3. The MoCA delayed recall score was not significantly correlated with the BVMT-R recognition discrimination index score. All other scores were significantly correlated with each other (ranging between $r = 0.10$, $p = 0.015$ and $r = 0.96$, $p < 0.001$). Specifically, scores from measures within tests (e.g., HVLT immediate recall and delayed recall) were strongly correlated with each other (ranging between $r = 0.39$, $p < 0.001$ and $r = 0.96$, $p < 0.001$), whereas scores between tests (e.g., HVLT

immediate recall and paired associates immediate recall) were only moderately correlated with each other (ranging between $r = 0.10$, $p = 0.015$ and $r = 0.52$, $p < 0.001$). Bartlett's test of sphericity was significant (χ^2 (120) = 8295.84, $p < 0.001$), indicating that a factor analytic approach was appropriate for reducing the dimensionality of the episodic memory test scores. The KMO measure of sampling adequacy indicated that the strength of the relationship among variables was high (KMO = 0.82) and, thus, it was acceptable to proceed with subsequent analyses.

The original proposed model with a general episodic memory domain factor and four subfactors (Figure 1) did not satisfy the model fit criteria (CFI = 0.941, RMSEA = 0.089, SRMR = 0.086, χ^2 (96) = 584.20, $p < 0.001$; Table 4 and 5). The standardized loadings of the four subfactors onto an episodic memory domain factor were not significant (ranging between 0.86 and 0.99, all $p = 0.98$). However, all test scores reliably loaded onto the four subfactors (absolute standardized loadings ranging between 0.35 and 0.75, all $p < 0.001$).

Due to the lack of a good fit for the original model, a series of alternative confirmatory factor analyses were examined. Based on available evidence and theoretical positions of episodic memory, a model that hypothesized a single episodic memory factor without any subfactors was tested. This model had a worse fit than the original model (CFI = 0.929, RMSEA = 0.095, SRMR = 0.182, χ^2 (100) = 686.64, $p < 0.001$; Table 6; Appendix B). Next, based on the alternative hypothesis outlined by several manuscripts, the hypothesis that a model without a general episodic memory factor and only subfactors would provide the best fit was tested. This model also showed worse fit than the original model (CFI = 0.940, RMSEA = 0.089, SRMR = 0.093, χ^2 (97) = 596.26, $p < 0.001$; Table 6; Appendix B).

Using the original model with a general episodic memory factor and subfactors, indicators with factor loadings less than 0.55 were removed (including all four Cohen's Relational Memory

Test scores and the two recognition discrimination index scores from the BVMT-R and HVLT-R). The model fit criteria were satisfied (CFI = 0.998, RMSEA = 0.025, SRMR = 0.018, χ^2 (28) = 38.93, p = 0.08; Figure 1, Table 5 and 6). Specifically, consistent with the original model and after removing indicators with low loadings, findings revealed an excellent model fit with three first-order subfactors derived from the second-order general episodic memory domain factor. The first-order subfactors were as follows: 1) a verbal immediate recall component derived from the Logical Memory total immediate recall score, Paired Associates immediate recall score, and HVLT-R total learning score; 2) a verbal delayed recall component derived from MoCA, Logical Memory, Paired Associates, and HVLT-R delayed recall scores; and 3) a visuospatial component derived from PSMT and BVMT-R total and delayed recall. All subfactors loaded reliably on the general episodic memory factor (standardized loadings ranging between 0.92 and 0.99, all $p < 0.001$). Further, all episodic memory scores loaded reliably on the three subfactors (standardized loadings ranging between 0.57 and 0.75, all $p < 0.001$). The post hoc power analysis revealed that a sample of N = 648 is associated with power larger than > 99.99% to reject a wrong model (df = 28) with an amount of misspecification corresponding to RMSEA = 0.08 and alpha = 0.05. All subsequent analyses used this model with three subfactors.

3.3 Hippocampal Volume

Seven participants were missing hippocampal volume data and 21 participants had a significant portion of the hippocampus clipped during image acquisition. Thus, all subsequent analyses include the remaining 620 participants. Independent samples *t*-tests revealed that participants with significantly clipped hippocampi did not differ significantly on key demographic characteristics from the rest of the sample, including age, gender, race, and education (all $p >$

0.066). There was one outlier for CA2 volume, but the value did not reflect measurement error and, thus, was kept in all subsequent analyses. Further, there were no missing values or outliers for the observed episodic memory factors. The observed factors and hippocampal volume variables were not heavily skewed and followed the normal distribution to an acceptable extent (absolute skewness range 0.02 – 0.58 and absolute kurtosis range 2.72 – 3.82) (Kline, 2015).

The *a priori* power analysis revealed that, in order to detect a Pearson's correlation coefficient of $r = 0.29$ with 80% power ($\alpha = 0.05$, two-tailed), a sample size of 91 participants is needed. Therefore, there was adequate power to detect a significant effect.

Correlations between demographic variables, the observed episodic memory factors, and hippocampal volumes are presented in Table 7. Collection site (Boston, Kansas City, Pittsburgh) was not significantly related with any variables of interest and, thus, was not included as a covariate in any models. Age, gender, race, and years of education were significantly correlated with the episodic memory factors (ranging between $r = 0.15$, $p < 0.001$ and $r = 0.23$, $p < 0.001$), such that younger age, female gender, White race, and higher education were associated with better episodic memory performance. Age, gender, race, and intracranial volume were significantly correlated with total, left, and right hippocampal volume (ranging between $r = 0.12$, $p = 0.002$ and $r = 0.36$, $p < 0.001$), such that younger age, male gender, White race, and greater intracranial volume were associated with greater hippocampal volume. Thus age, gender, race, education, and intracranial volume were included as covariates in all subsequent analyses.

Hierarchical regression models were used to test the hypothesis that the verbal delayed recall subfactor would explain the most variance in total, left, and right hippocampal volume, above and beyond the other subfactors. As depicted in Table 8, the regression analyses revealed that all subfactors were significantly associated with total, left, and right hippocampal volume

when considered independently in step two of the model after controlling for covariates (ranging between $\beta = 0.10$, $p = 0.016$ and $\beta = 0.14$, $p < 0.001$). The significant marginal effects of each subfactor were as follows: 1) the verbal immediate recall subfactor explained 1.20%, 1.43%, and 0.83% of the variance in total, left, and right hippocampal volume, respectively, 2) the verbal delayed recall subfactor explained 1.25%, 1.58%, and 0.80% of the variance in total, left, and right hippocampal volume, respectively, and 3) the visuospatial subfactor explained 1.19%, 1.49%, and 0.77% of the variance in total, left, and right hippocampal volume, respectively. Inconsistent with our hypothesis, all three subfactors explained a similar amount of variance in total, left, and right hippocampal volume. Adding all the subfactors in step three of the model did not significantly explain additional variance in total, left, or right hippocampal volume compared to each subfactor by itself (ranging from change in $R^2 < 0.001$, $p = 0.992$ and change in $R^2 = 0.002$, $p = 0.420$). When all the subfactors were included in step three of the model, none of the subfactors were significant predictors of total, left, or right hippocampal volume (ranging between $\beta = 0.31$, $p = 0.258$ and $\beta < 0.01$, $p = 0.984$). Together, the covariates and subfactors accounted for 21.8%, 21.0%, and 19.8% of the variance in total, left, and right hippocampal volume, respectively.

Hierarchical regression models were used to test whether a distinct subfactor explains significantly greater variation in hippocampal volume than the general domain factor of episodic memory. As depicted in Table 9, the results from these analyses revealed that the general episodic memory domain factor was significantly associated with total, left, and right hippocampal volume after controlling for covariates (ranging between $\beta = 0.10$, $p = 0.013$ and $\beta = 0.13$, $p < 0.001$). The general episodic memory domain factor uniquely explained 1.23%, 1.53%, and 0.82% of the variance in total, left, and right hippocampal volume, respectively. Inconsistent with our hypothesis, the general episodic memory domain factor explained a similar amount of variance in

total, left, and right hippocampal volume as all three subfactors. Together, the covariates and the general episodic memory domain factor accounted for 21.7%, 20.8%, and 19.7% of the variance in total, left, and right hippocampal volume, respectively.

3.4 Hippocampal Subfield Volume

Hierarchical regression models were used to test the hypothesis that the following hippocampal subfield volumes would vary in their involvement in episodic memory subfactors: 1) the ammonic subfields (CA1-3), 2) dentate gyrus, 3) entorhinal cortex, and 4) subiculum. As depicted in Table 10, all subfactors were significantly associated with CA1, entorhinal cortex, and subiculum volume when considered independently in step two of the model after controlling for covariates (ranging between $\beta = 0.114$, $p = 0.004$ and $\beta = 0.155$, $p < 0.001$). Only the verbal immediate recall and verbal delayed recall subfactors were significantly associated with CA3 volume in step two of the model after controlling for covariates ($\beta = 0.08$, $p = 0.041$ and $\beta = 0.09$, $p = 0.037$, respectively). Inconsistent with our hypothesis, none of the three subfactors were significantly associated with CA2 or dentate gyrus volume in step two of the model after controlling for covariates (ranging between $\beta = -0.03$, $p = 0.47$ and $\beta = 0.07$, $p = 0.09$). The significant marginal effects of each subfactor were as follows: 1) the verbal immediate recall subfactor explained 1.12%, 0.59%, 1.99%, and 1.40% of the variance in CA1, CA3, subiculum, and entorhinal cortex volume, respectively, 2) the verbal delayed recall subfactor explained 1.14%, 0.62%, 1.76%, and 1.37% of the variance in CA1, CA3, subiculum, and entorhinal cortex volume, respectively, and 3) the visuospatial subfactor explained 1.09%, 2.01%, and 1.27% of the variance in CA1, subiculum, and entorhinal cortex volume, respectively. All three subfactors explained a similar amount of variance in subfield volumes. However, adding all the subfactors in step three

of the model did not significantly explain additional variance in any of the subfield volumes compared to each subfactor by itself (ranging from change in $R^2 < 0.001$, $p = 0.99$ and change in $R^2 = 0.006$, $p = 0.10$). When all the subfactors were included in step three of the model, none of the subfactors were significant predictors of any of the subfield volumes (ranging between $\beta = -0.001$, $p = 0.99$ and $\beta = 0.46$, $p = 0.11$). Together, the covariates and subfactors accounted for 18.9%, 4.6%, 13.8%, 11.0%, 27.0%, and 27.8% of the variance in CA1, CA2, CA3, dentate gyrus, subiculum, and entorhinal cortex volume, respectively.

4.0 Discussion

Although there are various ways of conceptualizing and assessing episodic memory, previous research suggests that different materials and designs of episodic memory tasks are only moderately correlated with each other (Benedict et al., 1996; Shapiro et al., 1999; Sudo et al., 2019). Further, various episodic memory tasks exhibit disproportional task demands on the hippocampus and differentially reflect hippocampal volume degeneration (Sudo et al., 2019). Therefore, it is unclear if variation in performance on episodic memory measures is primarily due to variation in episodic memory ability or even whether it is a meaningful indicator of risk for developing Alzheimer's disease. This study established a structural equation model to examine the covariance structure and distinctiveness of tasks that have been traditionally used as measures of episodic memory and assessed whether these relate differently to hippocampal volume. Based on previous literature, it was predicted that there would be four subfactors (verbal immediate recall, verbal delayed recall, visuospatial, and recognition) derived from a general episodic memory domain factor, and the verbal delayed recall component would explain the most variance in hippocampal volume. Inconsistent with our hypothesis, a model with three subfactors (verbal immediate recall, verbal delayed recall, and visuospatial) derived from a general episodic memory domain factor had the best model fit. Although this model was not the original hypothesized model, it is still in line with current theories and conceptualizations of episodic memory. Further, all three subfactors and the general episodic memory domain factor explained a similar amount of variance in total, left, and right hippocampal volume. In addition, all subfactors were significantly associated with CA1, entorhinal cortex, and subiculum volume, only the verbal immediate recall and verbal delayed recall subfactors were significantly associated with CA3 volume, and none of the three subfactors were significantly associated with CA2 or dentate gyrus volume. However,

the amount of variance in hippocampal volume explained was minimal and similar across all subfactors. These results suggest that traditional episodic memory tasks are in fact measuring the same overarching construct, but different task conditions are tapping into different episodic memory processes. In addition, these findings indicate that examining multiple measures of episodic memory does not provide additional information than that obtained when examining only one measure or score of episodic memory. Further, this study suggests that different hippocampal subfields are not uniformly involved in managing and supporting episodic memory processes. Overall, the findings from this study indicate that hippocampal volume might not be a reliable marker of episodic memory performance among those without cognitive impairment.

4.1 Confirmatory Factor Analysis

Previous research has found that performance on a word list task, a story learning task, and a figure learning task were only moderately correlated with each other (Sudo et al., 2019). The results from this study are in line with previous studies and found that scores between tests were only moderately correlated with each other. These results suggest that traditional measures that are commonly and jointly referred to as 'episodic memory' tasks may be measuring several subtypes of episodic memory rather than a single, general, overarching construct. In line with these results, when conducting a factor analysis to determine whether the measures in this study represented one underlying construct or several subfactors of episodic memory, a model with no subfactors and only a general episodic memory domain factor had the worst fit. These results suggest that there is a general episodic memory construct with underlying processes that are more specific to task conditions and materials. The model slightly improved when four subfactors were included,

although it still did not satisfy the proposed model fit criteria. This suggests that not all episodic memory tasks included in this study are measuring the same constructs to a similar extent.

When removing scores that did not strongly load on the subfactors, the resulting model consisted of three subfactors and a significantly improved model fit. The scores that were removed included all four Cohen's Relational Memory Test scores and the two recognition discrimination index scores (BVMT-R and HVLT-R). The results regarding Cohen's Relational Memory Test are surprising, as an important component of episodic memory is relational memory (Eichenbaum & Cohen, 2004; Ngo et al., 2018) and the task has been found to be strongly tied to other episodic memory tasks and bilateral hippocampal volume (Monti et al., 2015). A follow-up sensitivity analysis was performed to assess the fit of a confirmatory factor analysis model with just the four Cohen's Relational Memory Test Scores and its association with hippocampal volume. The model had a poor model fit (CFI = 0.937, RMSEA = 0.391, SRMR = 0.065, χ^2 (2) = 199.87, $p < 0.001$; Appendix C) and the observed factor was not significantly associated with total, left, or right hippocampal volume after controlling for covariates (ranging between $\beta = 0.01$, $p = 0.75$ and $\beta = -0.003$, $p = 0.94$; Appendix C). Cohen's Relational Memory Test is different from the other tasks included in this study in several important ways: the delay was only 2 seconds as compared to a 5–25-minute delay, the elements presented were lines in locations as compared to semantically related words, shapes, or pictures, and the responses were made using a computer mouse to click and drag stimuli as compared to freely providing verbal or hand-drawn responses. It is possible that these important differences in the task design led to performance that is not dependent on the hippocampus. Future research is needed to assess whether scores on this task are more strongly associated with hippocampal function than volume. Studies would also benefit from examining whether the results differ when subdividing the hippocampus along the longitudinal axis (i.e., head,

body tail), as functional differences have been discovered between the anterior and posterior sections that are still poorly understood (Hrybouski et al., 2019). However, the results from this study suggest that Cohen's Relational Memory Test is measuring disparate cognitive processes more than a shared, underlying process with other episodic memory tasks.

The other two scores that were removed were the two recognition discrimination index scores from the BVMT-R and the HVLT-R. A previous principal components analysis found that the HVLT-R recognition discrimination index score loaded onto a factor with other HVLT-R scores, whereas the BVMT-R recognition discrimination index did not load significantly onto any factors, including a separate factor comprised of the other BVMT-R scores (Benedict et al., 1996). In contrast, another previous principal components factor analysis found that HVLT-R and BVMT-R recall and recognition discrimination index scores loaded onto one factor, and that performance on the HVLT-R recognition discrimination index was the most useful in discriminating between patients with Alzheimer's disease and vascular dementia (Shapiro et al., 1999). Both studies included patients with neuropsychiatric or neurodegenerative conditions. Given that the sample in this study consisted of older adults without clinically observable cognitive impairment, very few participants had low scores on the recognition portion of these tasks, yielding less variability in scores than studies with participants with impaired cognition. For example, while this study had a mean ± standard deviation of 10.62 ± 1.44 for HVLT recognition discrimination index scores, participants with an Alzheimer's disease diagnosis had a mean ± standard deviation of 5.1 ± 3.5 in a previous study (Shapiro et al., 1999). There was likely insufficient variability to detect subtle relationships with other episodic memory measures in this study. It is possible that the recognition discrimination index scores from the BVMT-R and the HVLT-R would have loaded more strongly in the models if the sample consisted of older adults with cognitive decline

and, thus, included more variability in performance. Therefore, these scores may become increasingly informative in later stages of pathological cognitive decline.

A confirmatory factor analysis investigated the subfactors of episodic memory since little work has explored the components derived from theory using a data-driven approach. The final, optimal model yielded three subfactors that were derived from the general episodic memory domain factor, indicating that these subfactors are likely representing their respective cognitive processes. The verbal immediate recall component was derived from the Logical Memory total immediate recall score, Paired Associates immediate recall score, and HVLT-R total learning score. The verbal delayed recall component was derived from the MoCA, Logical Memory, Paired Associates, and HVLT-R delayed recall scores. The visuospatial component was derived from the PSMT and BVMT-R total and delayed recall scores. These results suggest that the underlying cognitive processes of episodic memory are specific to task conditions (i.e., immediate vs delayed) and the presentation of material (i.e., verbal versus visuospatial), such that these scores should not be used interchangeably. These results corroborated the theory that episodic memory is mediated by complex processes of encoding and retrieval, and that these processes differ based on material type. Thus, the findings suggest that different episodic memory task conditions are tapping into these various processes. Given that the final model also included a general episodic memory domain factor, these results further suggest that traditional episodic memory tasks are in fact measuring the same overarching construct.

4.2 Hippocampal Volume

It is well known that the hippocampus is involved to some extent across all complex processes of episodic memory, including encoding, consolidation, storage, and retrieval (Aggleton

& Brown, 1999; Mayes, 1988). Longitudinal studies show that smaller hippocampal volume is one of the strongest predictors of rapid decline in episodic memory performance (Ottoy et al., 2019). Although episodic memory tasks exhibit disproportional demands on the hippocampus (Sudo et al., 2019), it remains unclear which measures are most sensitive to variation in hippocampal volume before clinically observable cognitive deficits are present. Hierarchical regression analyses revealed that all three subfactors and the general episodic memory domain factor were significantly associated with total, left, and right hippocampal volume. Further, the three subfactors and the general episodic memory domain factor explained a similar amount of variance in total, left, and right hippocampal volume (less than a 0.2% difference). These results contradict previous findings indicating that not all episodic memory tasks explain a significant amount of variance in hippocampal atrophy (Sudo et al., 2019). This discrepancy in findings might be due to differences in sample size, as the power analysis in this study determined they did not have adequate power to detect a significant effect. This discrepancy in findings might also be due to differences in sample characteristics, as their study consisted of older adults with and without cognitive impairment. The findings from this study suggest that verbal immediate recall, verbal delayed recall, and visuospatial performance are all informative of the structural integrity of the hippocampus among individuals without clinical memory impairment. Additionally, it suggests that examining multiple measures of episodic memory does not provide additional information than that obtained when examining only one measure or score of episodic memory. It is possible that there would be greater differences between the amount of variance explained by the subfactors as cognitive and structural brain deterioration becomes more apparent. It is also possible that the subfactors are more strongly correlated with in vivo measures of hippocampal function than structure.

It is important to note that, although the associations are statistically significant, all the explained variances are very low across all the subfactors and the general episodic memory domain factor (less than 2%). The amount of variance explained in this study is lower than in previous research (e.g., Sudo et al., (2019) 35-48%). One potential explanation for this difference is variations in methods used to normalize for head size. Sudo and colleagues (2019) used a calculated Hippocampal Occupancy Score (HOS) that takes into account hippocampal volume loss and the resulting increase in inferior lateral ventricle volume (CorTechs, 2020). However, the HOC score does not account for head size. While this study did not utilize a score of hippocampal atrophy, it did adjust for estimated intracranial volume. Given that people with larger heads typically have larger hippocampi, controlling for intracranial volume allowed us to assess the deviation of hippocampal volume from what would be expected given their head size. Removing variance in hippocampal volume associated with head size typically yields a smaller association between episodic memory performance and hippocampal volume (Van Petten, 2004). The small amount of variance explained in this study could potentially explain why many previous studies fail to find an association between episodic memory and hippocampal volume (Van Petten, 2004). Given that the observed results only account for a small portion of the variance in hippocampal volume, many other genetic and environmental factors are likely influencing hippocampal morphology. In fact, these results suggest that hippocampal volume might not be a reliable marker of episodic memory performance among those without cognitive impairment. Given that previous research suggests that smaller hippocampal volume at baseline is one of the strongest predictors of a faster decline in episodic memory and conversion to Alzheimer's disease (Ottoy et al., 2019), it is possible that the observed factors would explain greater variance in hippocampal volume later in the disease course. Future longitudinal studies can offer insight into the importance of

hippocampal volume for episodic memory performance as clinically observable cognitive deficits emerge.

4.3 Hippocampal Subfield Volume

The hippocampus is composed of various subfields, and there are several theories about the different functions of each and the degree to which atrophy progresses across the spectrum of pathological aging (Carlesimo et al., 2015; Mueller et al., 2011). However, it is unclear how the subfields vary in their involvement in distinct episodic memory processes and materials. Hierarchical regression analyses revealed that, while all subfactors were significantly associated with CA1, entorhinal cortex, and subiculum volume, only the verbal immediate recall and verbal delayed recall subfactors were significantly associated with CA3 volume, and none of the three subfactors were significantly associated with CA2 or dentate gyrus volume. The results regarding CA2 and dentate gyrus volume are surprising, as several previous studies have found both to be associated with verbal immediate and delayed recall performance (Aslaksen et al., 2018; Zheng et al., 2018), as well as delayed recall performance on visuospatial episodic memory tasks (Zammit et al., 2017). However, another study found that CA2 volume was not associated with verbal immediate or delayed recall performance and that dentate gyrus volume was only associated with verbal immediate recall performance (Mueller et al., 2011). This discrepancy in findings might be due to differences in sample characteristics and limited sample sizes. For example, studies that found an association included a small sample of adults across the lifespan, which may have added heterogeneity and reduced generalizability to this large sample of older adults (Aslaksen et al., 2018; Zheng et al., 2018). Further, the study by Mueller et al (2011) examined a small sample (N= 50) of older adults with and without cognitive impairment. As a result, some of their participants

may have Alzheimer's disease-related pathology that might have affected hippocampal-memory relationships. In addition, different tasks were employed across these studies, and, as displayed in this study, these tasks tend to only be moderately correlated with each other. Thus, without a factor analysis approach, different tasks may be measuring different aspects of the construct of episodic memory and may produce variability in hippocampal correlates.

Our results regarding significant associations of CA1, entorhinal cortex, and subiculum volumes with all three subfactors are somewhat in line with previous research. Specifically, while some studies have found associations between CA1 volume and only delayed verbal and visuospatial recall (Mueller et al., 2011; Zammit et al., 2017), another study found CA1 volume to also be associated with immediate verbal recall (Aslaksen et al., 2018). Further, entorhinal cortex volume has not been associated with immediate or delayed verbal episodic memory at baseline but a greater annual rate of entorhinal cortex shrinkage predicted worse performance over a 5-year period in one study of healthy adults (Rodrigue & Raz, 2004). Lastly, subiculum volume has been found to only be associated with delayed verbal and visuospatial recall (Zammit et al., 2017). The results from this study regarding significant associations of CA3 volume with the verbal immediate recall and verbal delayed recall subfactors are also somewhat in line with previous research. Specifically, CA3 volume has been found to only be associated with immediate verbal recall in one study (Mueller et al., 2011), while it was found to also be associated with delayed verbal recall in another study (Aslaksen et al., 2018). As mentioned above, this discrepancy in findings between this study and previous research might be due to differences in sample characteristics, tasks measured, and limited sample sizes. The results from this study also contradict current theories that argue that CA3 supports rapid learning and short-term retrieval, whereas CA1 recodes information from CA3 and allows for the retrieval of the information after longer time intervals

(Rolls & Kesner, 2006). They also differ from theorists that argue that the subiculum is important for processing spatial relations (O′Mara, 2005), whereas the entorhinal cortex is critical for spatial and object relations (Schultz et al., 2015). Overall, these results add muddiness to the literature and instead suggest that the volume of these subfields may not be purely responsible for any of these aspects of episodic memory processing. For example, these measures of episodic memory likely do not rely exclusively on visual or verbal processes, such that even visual measures may still recruit verbal strategies during encoding and retrieval. However, additional research is needed to corroborate these results.

It is also critical to note that, of the significant associations with hippocampal subfield volumes, all three subfactors explained a similar amount of variance (less than a 0.3% difference). These results contradict previous findings indicating specialization of hippocampal subfield volumes for performance across various episodic memory tasks and conditions (Aslaksen et al., 2018; Mueller et al., 2011). This discrepancy in findings might be due to differences in sample size, as the power analysis determined they did not have adequate power to detect a significant effect. This discrepancy in findings might also be due to differences in sample characteristics, as their studies consisted of adults across the lifespan and older adults with cognitive impairment. The findings from this study suggest that verbal immediate recall, verbal delayed recall, and visuospatial performance are all informative of the structural integrity of the hippocampal subfields among older adults without clinical memory impairment. Additionally, it suggests that examining multiple measures of episodic memory does not provide additional information than that obtained when examining only one measure or score of episodic memory. It is possible that there would be greater differences between the amount of variance explained by the subfactors as cognitive and structural brain deterioration becomes more apparent. It is also possible that the

subfactors are more strongly correlated with in vivo measures of hippocampal subfield function than structure.

Notably, although some of the associations are statistically significant, all the explained variances are very low across all the subfactors (less than 2.5%). The amount of variance explained in this study is lower than in previous research (e.g., Aslaksen et al., (2018) 16-22%). As mentioned previously, these differences may be due to variations in head size, sample size, characteristics of the sample, or episodic memory tasks used. This could potentially explain why some of the above studies failed to find an association between episodic memory and various hippocampal subfield volumes. Given that the observed results only account for a small portion of the variance in hippocampal subfield volume, many other genetic and environmental factors, such as physical activity, are likely influencing hippocampal morphology. In fact, these results suggest that hippocampal subfield volume might not be a strong marker of episodic memory performance before deficits emerge.

4.4 Limitations

Despite a well-characterized, large sample of older adults with multiple assessments of episodic memory, the current study was not without its limitations. First, given that this was a cognitively healthy sample, participants' episodic memory scores were well within the average range and may not have had sufficient variability towards the lower range of scores. Additional participants in an impaired range may have added more clinical relevance and shown different results. Specifically, it is possible that a wider range of scores would have transformed the associations between episodic memory and hippocampal subfield volume to be more in line with prior literature. Second, the sample consisted of healthy older adults without neurological

conditions, major psychiatric illnesses, substance abuse, or cardiovascular events or conditions, which may have equipped them with a number of protective factors that might reduce the risk of deteriorating health and limit the generalizability of the results. Third, this study examined baseline data, which limits our ability to draw conclusions about how performance on episodic memory tasks predicts future decline or changes in hippocampal morphology. Fourth, participants in the IGNITE study completed their neuropsychological assessments over the course of two days, which may have affected performance. While cognitive abilities are thought to be stable in the short-term, neuropsychological testing on separate days may have introduced potential confounds, such as fluctuations in attention and fatigue. However, variations in attention and energy are common across all cognitive testing regardless of duration; thus, these results are likely generalizable to everyday cognition. Fifth, the episodic memory tasks examined in this study represent complex memory processes that involve attention and executive functions and, thus, do not rely exclusively on hippocampal-related functions. While this study does not examine other cognitive non-episodic processes or related brain regions, future work would benefit from extending these results to other cognitive domains and brain areas.

4.5 Contributions

Despite these limitations, there are several strengths of the current study. First, there are several advantages to this sample. This study consisted of a large sample, which provided greater power than previous studies to conduct a factor analysis (e.g., $N_{Shaprio(1999)} = 445$, $N_{Benedict(1996)} = 457$) and examine the association between episodic memory and hippocampal volume (e.g., $N_{Sudo(2019)} = 27$, $N_{VanPetten(2004)} = 48$). In addition to a large sample, this study focused on healthy older adults, whereas previous research examined adults across the lifespan and various disease

states. Thus, these prior studies may have increased variability and hindered their ability to detect a significant effect. In contrast, this study consisted of a well-characterized sample of the general aging population and is more generalizable to late adulthood. This sample also included greater racial and ethnic diversity (75% White) than previous studies in this area (e.g., Benedict et al., (1996) 80% White). Differences in racial and ethnic makeup may have led to discrepancies in the findings observed in this study compared to previous studies. However, it is important to note that while the racial distribution in this study is representative of the cities in which recruitment occurred, the racial and ethnic composition of this study is not characteristic of the general United States population, restricting the generalizability of these results. Lastly, this sample consisted of physically inactive older adults, whereas previous studies did not screen for activity levels. Given that it is estimated that 67% of the older adult population is sedentary (Harvey et al., 2013), the results from this study may be more characteristic of the general aging population than previous studies.

There were also several strengths related to the analyses. First, episodic memory was comprehensively measured using seven of the most commonly used tasks, which permitted us to examine their covariance structure and distinctiveness. Second, using a factor analysis approach allowed us to comprehensively capture the different processes of episodic memory (i.e., encoding, retrieval) within each task and reduce measurement error while accounting for the shared variance. Lastly, this study used advanced neuroimaging techniques to assess the volume of the hippocampus and its subfields.

This study examined associations between episodic memory and hippocampal volume in late adulthood to better understand which episodic memory measures are most reliable and predictive of a critical marker of future decline – hippocampal volume. Overall, the results suggest

that traditional episodic memory measures are in fact measuring the same overarching construct, but that not all tasks are measuring the same process to a similar extent. Instead, the underlying cognitive processes of episodic memory are specific to task conditions (i.e., immediate vs delayed) and the presentation of material (i.e., verbal versus visuospatial). Further, these various episodic memory processes explained a similar amount of variance in hippocampal volume, suggesting that examining multiple measures of episodic memory does not provide additional benefit than examining only one measure or score of episodic memory. This study also suggests that while various episodic memory measures provide quick, sensitive insight into the general domain of episodic memory, it is unclear whether performance is a meaningful indicator of the structural integrity of the hippocampus or its subfields. These results have wide-reaching implications for clinical neuropsychologists, neurologists, and researchers, as it suggests that performance on episodic memory measures should be corroborated with measures of brain health to accurately inform diagnoses and treatment recommendations. Further, clinicians and researchers would benefit from being selective in their measures when assessing episodic memory, as not all measures provide similar insight into various episodic memory processes. These results allow for the development of more precise neuropsychological protocols for detecting subtle changes in episodic memory among older adults with healthy cognition.

4.6 Future Directions

Given the cross-sectional nature of this study, longitudinal studies are needed to investigate several important questions. First, longitudinal studies would allow us to further our understanding of the statistical associations between episodic memory performance, hippocampal volume, and incidences of Alzheimer's disease in causal terms. Specifically, it will be imperative to examine

whether baseline performance among the subfactors predicts the development of Alzheimer's disease. Further, it will help determine whether a similar factor structure holds across the spectrum of pathological aging. Additional work is also needed to determine how these different episodic memory processes relate to hippocampal volume as pathology increases. It is possible that other subfactors become of critical importance in explaining variance in hippocampal volume as cognitive impairment advances. Lastly, it would highlight how strongly hippocampal volume at baseline predicts the development of pathology.

Future research is also needed to assess the hippocampus, other brain regions, and other non-episodic cognitive processes more comprehensively. Specifically, more research is needed to determine whether the subfactors are more strongly correlated with in vivo measures of hippocampal function than structure. Given that hippocampal volume declines at an accelerated rate after age 50 (Fjell et al., 2013), additional research is needed to determine how our results might vary across the lifespan using multilevel, and non-linear, models. In addition, future studies are necessary to examine how other cognitive non-episodic processes, such as attention, language, and executive functions, affected episodic memory abilities and related to the observed subfactors in this study. Further, the contribution of other brain regions, such as frontal neocortical regions, needs to be assessed in order to examine the impact they have on episodic memory processes, such as encoding and retrieval. Lastly, given that the observed results only account for a small portion of the variance in hippocampal volume, the link between hippocampal morphology and episodic memory performance is tenuous at best. Future research is needed to examine other genetic and environmental factors that more strongly influence hippocampal morphology, such as physical activity.

4.7 Conclusions

This study examined associations between episodic memory and hippocampal volume in late adulthood to better understand which episodic memory measures are most reliable and predictive of future decline. This study largely suggests that, while many widely used instruments designed for measuring episodic memory are in fact measuring the same overarching construct, they are not equivalently assessing the same process that is generally referred to as episodic memory. This study provides evidence that there are underlying processes that are more specific to task conditions and materials. Further, verbal delayed recall scores, verbal immediate recall scores, and visuospatial scores across multiple episodic memory measures are similarly linked with hippocampal volume, but they only account for a small portion of the variance in hippocampal volume. This suggests that the link between hippocampal morphology and episodic memory performance is questionable. This study also provides preliminary evidence that hippocampal subfields may not be purely responsible for any one aspect of episodic memory processing, but they may be preferentially important for various processes. However, this study adds muddiness to the literature and brings into question whether hippocampal subfield volume is a strong marker of episodic memory performance. The findings generated by this study of cognitively normal older adults lay the groundwork for determining which tasks and scores across episodic memory measures are most correlated with a critical Alzheimer's disease biomarker before a decline in cognition is clinically detectible. Future research would benefit from examining these results longitudinally to ascertain whether the factor structure of episodic memory and its relationship with hippocampal volume shifts across the spectrum of pathological aging.

5.0 References

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *The Behavioral and Brain Sciences*, *22*(3), 425–444; discussion 444-489.

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & 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: The Journal of the Alzheimer's Association*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Aly, M., & Turk-Browne, N. B. (2017). How hippocampal memory shapes, and is shaped by, attention. In *The hippocampus from cells to systems: Structure, connectivity, and functional contributions to memory and flexible cognition* (pp. 369–403). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-50406-3_12

Aslaksen, P. M., Bystad, M. K., Ørbo, M. C., & Vangberg, T. R. (2018). The relation of hippocampal subfield volumes to verbal episodic memory measured by the California Verbal Learning Test II in healthy adults. *Behavioural Brain Research*, *351*, 131–137. https://doi.org/10.1016/j.bbr.2018.06.008

Bäckman, L., Jones, S., Berger, A.-K., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology*, *19*(4), 520–531. https://doi.org/10.1037/0894-4105.19.4.520

Benedict, R. H. B., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test—Revised: Normative data and analysis of inter-form and test–retest reliability. *Clinical Neuropsychologist*, *12*(1), 43–55. https://doi.org/10.1076/clin.12.1.43.1726

Benedict, R. H. B., 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. https://doi.org/10.1037/1040-3590.8.2.145

Bigelow, J., & Poremba, A. (2014). Achilles' Ear? Inferior Human Short-Term and Recognition Memory in the Auditory Modality. *PLoS ONE*, *9*(2), e89914. https://doi.org/10.1371/journal.pone.0089914

Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *Clinical Neuropsychologist*, *5*(2), 125–142. https://doi.org/10.1080/13854049108403297

Brandt, J., & Benedict, R. H. B. (2001). *Hopkins Verbal Learning Test—Revised. Professional manual*. Psychological Assessment Resources, Inc.

Buckner, R. L., Head, D., Parker, J., Fotenos, A. F., Marcus, D., Morris, J. C., & Snyder, A. Z. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: Reliability and validation against manual measurement of total intracranial volume. *NeuroImage*, *23*(2), 724–738. https://doi.org/10.1016/j.neuroimage.2004.06.018

Carlesimo, G. A., Piras, F., Orfei, M. D., Iorio, M., Caltagirone, C., & Spalletta, G. (2015). Atrophy of presubiculum and subiculum is the earliest hippocampal anatomical marker of

Alzheimer's disease. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, *1*(1), 24–32. https://doi.org/10.1016/j.dadm.2014.12.001

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis, 2nd ed* (pp. xii, 430). Lawrence Erlbaum Associates, Inc.

CorTechs. (2020). Predicting Progression of MCI: Hippocampal Occupancy Score (HOC). *Cortechs.Ai*. https://www.cortechs.ai/predicting-progression-of-mci/

D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g1. *Biometrika*, *57*(3), 679–681. https://doi.org/10.1093/biomet/57.3.679

Dikmen, S. S., Bauer, P. J., Weintraub, S., Mungas, D., Slotkin, J., Beaumont, J. L., Gershon, R., Temkin, N. R., & Heaton, R. K. (2014). Measuring episodic memory across the lifespan: NIH Toolbox Picture Sequence Memory Test. *Journal of the International Neuropsychological Society: JINS*, *20*(6), 611–619. https://doi.org/10.1017/S1355617714000460

DiNapoli, E. A., Wu, B., & Scogin, F. (2014). Social Isolation and Cognitive Function in Appalachian Older Adults. *Research on Aging*, *36*(2), 161–179. https://doi.org/10.1177/0164027512470704

Eichenbaum, H., & Cohen, N. J. (2004). *From Conditioning to Conscious Recollection: Memory systems of the brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195178043.001.0001

Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, *13*(3), 230–253. https://doi.org/10.1037/a0013219

Erickson, K., Grove, G. A., Burns, J. M., Hillman, C. H., Kramer, A. F., McAuley, E., Vidoni, E. D., Becker, J. T., Butters, M. A., Gray, K., Huang, H., Jakicic, J. M., Kamboh, M. I., Kang, C., Klunk, W. E., Lee, P., Marsland, A. L., Mettenburg, J., Rogers, R. J., … Wollam, M. E. (2019). Investigating Gains in Neurocognition in an Intervention Trial of Exercise (IGNITE): Protocol. *Contemporary Clinical Trials*, *85*, 105832. https://doi.org/10.1016/j.cct.2019.105832

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fjell, A. M., Westlye, L. T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Holland, D., Dale, A. M., Walhovd, K. B., & Alzheimer Disease Neuroimaging Initiative. (2013). Critical ages in the life course of the adult brain: Nonlinear subcortical aging. *Neurobiology of Aging*, *34*(10), 2239–2247. https://doi.org/10.1016/j.neurobiolaging.2013.04.006

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews. Neurology*, *6*(2), 67–77. https://doi.org/10.1038/nrneurol.2009.215

Gaffan, D., & Hornak, J. (1997). Amnesia and neglect: Beyond the Delay-Brion system and the Hebb synapse. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1360), 1481. https://doi.org/10.1098/rstb.1997.0135

Groemping, U. (2007). Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software*, *17*, 1–27. https://doi.org/10.18637/jss.v017.i01

Harvey, J. A., Chastin, S. F. M., & Skelton, D. A. (2013). Prevalence of Sedentary Behavior in Older Adults: A Systematic Review. *International Journal of Environmental Research and Public Health*, *10*(12), 6645–6661. https://doi.org/10.3390/ijerph10126645

He, W., Goodkind, D., & Kowal, P. (2016). *An Aging World: 2015* (International Population Reports P95/16-1; Census Bureau). U.S. Government Publishing Office.

Hill, B. D., Alosco, M., Bauer, L., & Tremont, G. (2012). The relation of executive functioning to CVLT-II learning, memory, and process indexes. *Applied Neuropsychology. Adult*, *19*(3), 198–206. https://doi.org/10.1080/09084282.2011.643960

Hodges, J., Tulving, E., & Craik, F. (2000). Memory in the dementias. In *The Oxford Handbook of Memory* (pp. 441–459). Oxford University Press.

Hrybouski, S., MacGillivray, M., Huang, Y., Madan, C. R., Carter, R., Seres, P., & Malykhin, N. V. (2019). Involvement of hippocampal subfields and anterior-posterior subregions in encoding and retrieval of item, spatial, and associative memories: Longitudinal versus transverse axis. *NeuroImage*, *191*, 568–586. https://doi.org/10.1016/j.neuroimage.2019.01.061

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*(2), 139–154. https://doi.org/10.1037/0096-3445.122.2.139

Jeneson, A., Mauldin, K. N., & Squire, L. R. (2010). Intact working memory for relational information after medial temporal lobe damage. *The Journal of Neuroscience: The Official*

Journal of the Society for Neuroscience, *30*(41), 13624–13629. https://doi.org/10.1523/JNEUROSCI.2895-10.2010

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kapur, S., Craik, F. I., Tulving, E., Wilson, A. A., Houle, S., & Brown, G. M. (1994). Neuroanatomical correlates of encoding in episodic memory: Levels of processing effect. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(6), 2008–2011.

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (Fourth edition). The Guilford Press.

Langston, R. F., Stevenson, C. H., Wilson, C. L., Saunders, I., & Wood, E. R. (2010). The role of hippocampal subregions in memory for stimulus associations. *Behavioural Brain Research*, *215*(2), 275–291. https://doi.org/10.1016/j.bbr.2010.07.006

Lisman, J. E., & Otmakhova, N. A. (2001). Storage, recall, and novelty detection of sequences by the hippocampus: Elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. *Hippocampus*, *11*(5), 551–568. https://doi.org/10.1002/hipo.1071

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Madan, C. R. (2020). Rethinking the definition of episodic memory. *Canadian Journal of Experimental Psychology = Revue Canadienne De Psychologie Experimentale*, *74*(3), 183–192. https://doi.org/10.1037/cep0000229

Mayes, A. R. (1988). *Human organic memory disorders* (pp. vii, 300). Cambridge University Press. https://doi.org/10.1017/CBO9780511574597

Mayes, A. R., & Montaldi, D. (1999). The neuroimaging of long-term memory encoding processes. *Memory (Hove, England)*, *7*(5–6), 613–659. https://doi.org/10.1080/096582199387788

Mayes, A. R., & Roberts, N. (2001). Theories of episodic memory. *Philosophical Transactions of the Royal Society of London. Series B*, *356*(1413), 1395–1408. https://doi.org/10.1098/rstb.2001.0941

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011). The diagnosis of dementia 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: The Journal of the Alzheimer's Association*, *7*(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Millsap, R. E. (2001). When trivial constraints are not trivial: The choice of uniqueness constraints in confirmatory factor analysis. *Structural Equation Modeling*, *8*(1), 1–17. https://doi.org/10.1207/S15328007SEM0801_1

Monti, J. M., Cooke, G. E., Watson, P. D., Voss, M. W., Kramer, A. F., & Cohen, N. J. (2015). Relating hippocampus to relational memory processing across domains and delays. *Journal of Cognitive Neuroscience*, *27*(2), 234–245. https://doi.org/10.1162/jocn_a_00717

Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling*, *23*, 54–60. https://doi.org/10.1080/10705511.2014.950896

Mueller, S. G., Chao, L. L., Berman, B., & Weiner, M. W. (2011). Evidence for functional specialization of hippocampal subfields detected by MR subfield volumetry on high resolution images at 4 T. *NeuroImage*, *56*(3), 851–857. https://doi.org/10.1016/j.neuroimage.2011.03.028

Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Sun, L. D., Kato, A., Carr, C. A., Johnston, D., Wilson, M. A., & Tonegawa, S. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science (New York, N.Y.)*, *297*(5579), 211–218. https://doi.org/10.1126/science.1071795

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*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Ngo, C. T., Newcombe, N. S., & Olson, I. R. (2018). The Ontogeny of Relational Memory and Pattern Separation. *Developmental Science*, *21*(2), 10.1111/desc.12556. https://doi.org/10.1111/desc.12556

O'Mara, S. (2005). The subiculum: What it does, what it might do, and what neuroanatomy has yet to tell us. *Journal of Anatomy*, *207*(3), 271–282. https://doi.org/10.1111/j.1469-7580.2005.00446.x

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311–345. https://doi.org/10.1037/0033-295x.108.2.311

Ottoy, J., Niemantsverdriet, E., Verhaeghe, J., De Roeck, E., Struyfs, H., Somers, C., Wyffels, L., Ceyssens, S., Van Mossevelde, S., Van den Bossche, T., Van Broeckhoven, C., Ribbens,

A., Bjerke, M., Stroobants, S., Engelborghs, S., & Staelens, S. (2019). Association of short-term cognitive decline and MCI-to-AD dementia conversion with CSF, MRI, amyloid- and 18F-FDG-PET imaging. *NeuroImage. Clinical*, *22*, 101771. https://doi.org/10.1016/j.nicl.2019.101771

Petersen, R. C., Lopez, O., Armstrong, M. J., Getchius, T. S. D., Ganguli, M., Gloss, D., Gronseth, G. S., Marson, D., Pringsheim, T., Day, G. S., Sager, M., Stevens, J., & Rae-Grant, A. (2018). Practice guideline update summary: Mild cognitive impairment. *Neurology*, *90*(3), 126–135. https://doi.org/10.1212/WNL.0000000000004826

Rajan, K. B., Weuve, J., Barnes, L. L., McAninch, E. A., Wilson, R. S., & Evans, D. A. (2021). Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020-2060). *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*. https://doi.org/10.1002/alz.12362

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., & Acker, J. D. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex (New York, N.Y.: 1991)*, *15*(11), 1676–1689. https://doi.org/10.1093/cercor/bhi044

Rodrigue, K. M., & Raz, N. (2004). Shrinkage of the Entorhinal Cortex over Five Years Predicts Memory Performance in Healthy Adults. *The Journal of Neuroscience*, *24*(4), 956–963. https://doi.org/10.1523/JNEUROSCI.4166-03.2004

Rolls, E. T., & Kesner, R. P. (2006). A computational theory of hippocampal function, and empirical tests of the theory. *Progress in Neurobiology*, *79*(1), 1–48. https://doi.org/10.1016/j.pneurobio.2006.04.005

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. (2014). Correlates of cognitive change. *Journal of Experimental Psychology. General*, *143*(3), 1026–1048. https://doi.org/10.1037/a0034847

Salthouse, T. A., Fristoe, N., & Rhee, S. H. (1996). How localized are age-related effects on neuropsychological measures? *Neuropsychology*, *10*(2), 272–285. https://doi.org/10.1037/0894-4105.10.2.272

Schultz, H., Sommer, T., & Peters, J. (2015). The Role of the Human Entorhinal Cortex in a Representational Account of Memory. *Frontiers in Human Neuroscience*, *9*, 628. https://doi.org/10.3389/fnhum.2015.00628

Shapiro, A. M., Benedict, R. H., Schretlen, D., & Brandt, J. (1999). Construct and concurrent validity of the Hopkins Verbal Learning Test-revised. *The Clinical Neuropsychologist*, *13*(3), 348–358. https://doi.org/10.1076/clin.13.3.348.1749

Shi, F., Liu, B., Zhou, Y., Yu, C., & Jiang, T. (2009). Hippocampal volume and asymmetry in mild cognitive impairment and Alzheimer's disease: Meta-analyses of MRI studies. *Hippocampus*, *19*(11), 1055–1064. https://doi.org/10.1002/hipo.20573

Squire, L. R. (1986). Mechanisms of memory. *Science (New York, N.Y.)*, *232*(4758), 1612–1619. https://doi.org/10.1126/science.3086978

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: A neurobiological perspective. *Current Opinion in Neurobiology*, *5*(2), 169–177. https://doi.org/10.1016/0959-4388(95)80023-9

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021766. https://doi.org/10.1101/cshperspect.a021766

Sudo, F. K., de Souza, A. S., Drummond, C., Assuncao, N., Teldeschi, A., Oliveira, N., Rodrigues, F., Santiago-Bravo, G., Calil, V., Lima, G., Erthal, P., Bernardes, G., Monteiro, M., Tovar-Moll, F., & Mattos, P. (2019). Inter-method and anatomical correlates of episodic memory tests in the Alzheimer's Disease spectrum. *PloS One*, *14*(10), e0223731. https://doi.org/10.1371/journal.pone.0223731

Tomaszewski Farias, S., Cahn-Weiner, D. A., Harvey, D. J., Reed, B. R., Mungas, D., Kramer, J. H., & Chui, H. (2009). Longitudinal changes in memory and executive functioning are associated with longitudinal change in instrumental activities of daily living in older adults. *The Clinical Neuropsychologist*, *23*(3), 446–461. https://doi.org/10.1080/13854040802360558

Tremont, G., Halpert, S., Javorsky, D. J., & Stern, R. A. (2000). Differential impact of executive dysfunction on verbal list learning and story recall. *The Clinical Neuropsychologist*, *14*(3), 295–302. https://doi.org/10.1076/1385-4046(200008)14:3;1-P;FT295

Tulving, E. (1995). Organization of memory: Quo vadis? In *The cognitive neurosciences* (pp. 839–853). The MIT Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General*, *107*(4), 393–419. https://doi.org/10.1037/0096-3445.107.4.393

Van Petten, C. (2004). Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: Review and meta-analysis. *Neuropsychologia*, *42*(10), 1394–1413. https://doi.org/10.1016/j.neuropsychologia.2004.04.006

Van Petten, C., Plante, E., Davidson, P. S. R., Kuo, T. Y., Bajuscak, L., & Glisky, E. L. (2004). Memory and executive function in older adults: Relationships with temporal and prefrontal gray matter volumes and white matter hyperintensities. *Neuropsychologia*, *42*(10), 1313–1335. https://doi.org/10.1016/j.neuropsychologia.2004.02.009

Watson, P. D., Voss, J. L., Warren, D. E., Tranel, D., & Cohen, N. J. (2013). Spatial reconstruction by patients with hippocampal damage is dominated by relational memory errors. *Hippocampus*, *23*(7), 570–580. https://doi.org/10.1002/hipo.22115

Wechsler, D. (1997). *Wechsler Memory Scale* (3rd ed.). The Psychological Corporation.

Yee, L. T. S., Hannula, D. E., Tranel, D., & Cohen, N. J. (2014). Short-term retention of relational memory in amnesia revisited: Accurate performance depends on hippocampal integrity. *Frontiers in Human Neuroscience*, *8*, 16. https://doi.org/10.3389/fnhum.2014.00016

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews Neuroscience*, *20*(6), Article 6. https://doi.org/10.1038/s41583-019-0150-4

Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S.-L., Gertje, E. C., Mancuso, L., Kliot, D., Das, S. R., & Wolk, D. A. (2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapping*, *36*(1), 258–287. https://doi.org/10.1002/hbm.22627

Zammit, A. R., Ezzati, A., Zimmerman, M. E., Lipton, R. B., Lipton, M. L., & Katz, M. J. (2017). Roles of hippocampal subfields in verbal and visual episodic memory. *Behavioural Brain Research*, *317*, 157–162. https://doi.org/10.1016/j.bbr.2016.09.038

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, *78*(4), 16–33. https://doi.org/10.1111/mono.12032

Zheng, F., Cui, D., Zhang, L., Zhang, S., Zhao, Y., Liu, X., Liu, C., Li, Z., Zhang, D., Shi, L., Liu, Z., Hou, K., Lu, W., Yin, T., & Qiu, J. (2018). The Volume of Hippocampal Subfields in Relation to Decline of Memory Recall Across the Adult Lifespan. *Frontiers in Aging Neuroscience*, *10*. https://www.frontiersin.org/article/10.3389/fnagi.2018.00320

Appendix A

Table 1. Episodic Memory Measures and Conditions Included in the Hypothesized and Final Factor Analysis Models

Measure	Condition	Hypothesized	Final
Montreal Cognitive Assessment			
	Delayed free recall	✔	✔
Logical Memory			
	Total immediate free recall	✔	✔
	Delayed free recall	✔	✔
Paired Associates			
	Immediate free recall	✔	✔
	Delayed free recall	✔	✔
Hopkins Verbal Learning Test-Revised			
	Total learning free recall	✔	✔
	Delayed free recall	✔	✔
	Recognition discrimination index	✔	
Picture Sequence Memory Test			
	Cumulative number of pairs	✔	✔
Brief Visuospatial Memory Test-Revised			
	Total free recall	✔	✔
	Delayed free recall	✔	✔
	Recognition discrimination index	✔	
Cohen's Relational Memory Test			
	Misplacement	✔	
	Edge resizing	✔	
	Distortion	✔	
	Swaps	✔	

Table 2. Sample and Episodic Memory Performance Characteristics (N = 648)

Variable	Raw Mean (± SD)*	Normed Mean (± SD)+
Age (years)	69.88 (±3.75)	-
Gender (% female)	71.1	-
Education (years)	16.32 (±2.21)	-
Race (% White)	75.8	-
Montreal Cognitive Assessment Delay	3.02 (±1.55)	-
Logical Memory Total Immediate	43.43 (±9.03)	-
Logical Memory Delay	27.44 (±7.01)	-
Paired Associates Immediate	2.12 (±1.41)	-
Paired Associates Delay	1.43 (±1.39)	-
HVLT Total Immediate	26.00 (±4.49)	53.10 (±9.00)
HVLT Delay	9.15 (±2.11)	51.74 (±8.65)
HVLT Recognition Discrimination Index	10.62 (±1.44)	51.40 (±8.22)
Picture Sequencing Memory Test	10.37 (±5.93)	50.42 (±9.94)
Cohen's Relational Memory Test Misplacement	327.26 (±50.82)	-
Cohen's Relational Memory Test Edge Resize	494.03 (±81.40)	-
Cohen's Relational Memory Test Distortion	0.38 (±0.06)	-
Cohen's Relational Memory Test Swaps	0.15 (±0.05)	-
BVMT Total Immediate	21.10 (±6.42)	50.13 (±11.17)
BVMT Delay	8.66 (±2.53)	52.87 (±10.80)
BVMT Recognition Discrimination Index	5.63 (±0.70)	-

***** Raw scores were used in factor analyses

+ Normed scores reflect t-scores

All values (except gender and race) represent means ± standard deviations.

Table 3. Pearson's Correlations Between Episodic Memory Measures

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. MoCA_D	-	.36*	.39**	35**	.38**	.40**	.45**	.38**	.27**	-.19**	-.18**	-.14**	-.18**	.34**	r=.31**	.08
2. LM_I		-	.87**	.41**	.41**	.50**	.50**	.40**	.34**	-.30**	-.28**	-.26**	-.24**	.40**	r=.41**	.20**
3. LM_D			-	.38**	.40**	.47**	.52**	.43**	.37**	-.26**	-.24**	-.23**	-.21**	.39**	r=.40**	.17**
4. PA_I				-	.89**	.40**	.39**	.30**	.35**	-.33**	-.34**	-.29**	-.28**	.31**	r=.32**	.10*
5. PA_D					-	.41**	.41**	.31**	.39**	-.34**	-.35**	-.27**	-.30**	.31**	r=.33**	.10*
6. HVLT_I						-	.78**	.53**	.41**	-.30**	-.28**	-.30**	-.20**	.43**	r=.42**	.12**
7. HVLT_D							-	.63**	.40**	-.26**	-.24**	-.24**	-.19**	.44**	r=.46**	.12**
8. HVLT_R								-	.31**	-.18**	-.16**	-.18**	-.14**	.35**	r=.36**	.14**
9. PSMT									-	-.35**	-.34**	-.29**	-.29**	.39**	r=.38**	.18**
10. CRMT_M										-	.96**	.71**	.81**	-.38**	r=-.37**	-.13**
11. CRMT_E											-	.70**	.84**	-.34**	r=-.33**	-.12**
12. CRMT_D												-	.39**	-.30**	r=-.32**	-.11**
13. CRMT_S													-	-.30**	r=-.29**	-.11**
14. BVMT_I														-	r=.87**	.39**
15. BVMT_D															-	.42**
16. BVMT_R																-

* Correlation is significant at the p<.05 level

** Correlation is significant at the p<.01 level

Notes: Values reflect Pearson's correlation coefficient (r). MoCA, Montreal Cognitive Assessment Delay; LM, Logical Memory; PA, Paired Associates; HVLT, Hopkins Verbal Learning Test; PSMT, Picture Sequence Memory Test; CRMT, Cohen's Relational Memory Test; Brief Visuospatial Memory Test; I, immediate; D, delay; R, recognition; M, misplacement; E, edge resize; Di, distortion; S, swaps

Table 4. Factor Loadings Derived from Hypothesized Confirmatory Factor Analysis Model

2nd Order Factor	1st Order Factor	Loading	Measure	Loading
Episodic Memory	Verbal Immediate Recall	0.98		
			Logical Memory Immediate	0.69**
			Paired Associates Immediate	0.58**
			Hopkins Verbal Learning Test Immediate	0.69**
	Verbal Delayed Recall	0.99		
			Montreal Cognitive Assessment Delay	0.57**
			Logical Memory Delay	0.68**
			Paired Associates Delay	0.60**
			Hopkins Verbal Learning Test Delay	0.73**
	Visuospatial	0.86		
			Picture Sequence Memory Test	0.58**
			Brief Visuospatial Memory Test Immediate	0.75**
			Brief Visuospatial Memory Test Delay	0.75**
			Cohen's Relational Memory Test Misplacement	-0.52**
			Cohen's Relational Memory Test Edge Resize	-0.50**
			Cohen's Relational Memory Test Distortion	-0.35**
			Cohen's Relational Memory Test Swaps	-0.38**
	Recognition	0.94		
			Hopkins Verbal Learning Test Recognition	0.43**
			Brief Visuospatial Memory Test Recognition	0.39**

****** Significant at the p<.01 level

Table 5. Factor Loadings Derived from the Final Confirmatory Factor Analysis Model

2nd Order Factor	1st Order Factor	Loading	Measure	Loading
Episodic Memory	Verbal Immediate Recall	0.98**		
			Logical Memory Immediate	0.68**
			Paired Associates Immediate	0.57**
			Hopkins Verbal Learning Test Immediate	0.73**
	Verbal Delayed Recall	0.99**		
			Montreal Cognitive Assessment Delay	0.58**
			Logical Memory Delay	0.69**
			Paired Associates Delay	0.59**
			Hopkins Verbal Learning Test Delay	0.75**
	Visuospatial	0.92**		
			Picture Sequence Memory Test	0.60**
			Brief Visuospatial Memory Test Immediate	0.64**
			Brief Visuospatial Memory Test Delay	0.64**

***** Significant at the p<.05 level
****** Significant at the p<.01 level

Table 6. Model Fit Indices

Model	χ^2	df	χ^2 sig.	CFI	RMSEA	SRMR
Hypothesized	584.20	96	<0.001	0.941	0.089	0.086
Hypothesized without subfactors	686.64	100	<0.001	0.929	0.095	0.182
Hypothesized without general domain factor	596.26	97	<0.001	0.940	0.089	0.093
Final	38.93	28	0.08	0.998	0.025	0.018

Table 7. Pearson's Correlations Between All Observed Variables of Interest

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Age	-	-.12**	-.02	.09*	.12**	.06	-.19**	-.18**	-.21**	-.19**	-.24**	-.25**	-.22**
2. Gender		-	-.14**	-.13**	-.03	-.55**	.21**	.21**	.20**	.21**	-.17**	-.18**	-.16**
3. Race			-	.18**	.08*	.18**	.15**	.16**	.17**	.16**	.13**	.12**	.13**
4. Education				-	.07	.21**	.23**	.22**	.22**	.22**	.06	.04	.07
5. Site					-	.04	-.06	-.05	-.07	-.06	-.03	-.02	-.04
6. ICV						-	-.04	-.06	-.03	-.05	.36**	.34**	.36**
7. VIR							-	.99**	.98**	.99**	.14**	.15**	.13**
8. VDR								-	.98**	.99**	.14**	.15**	.12**
9. VIS									-	.99**	.15**	.16**	.14**
10. EM										-	.14**	.15**	.13**
11. Total HV											-	.96**	.96**
12. Left HV												-	.86**
13. Right HV													-

* Correlation is significant at the p<.05 level

** Correlation is significant at the p<.01 level

Notes: Values reflect Pearson's correlation coefficient (r). Gender = male: 1, female: 2; ICV, intracranial volume; VIR, verbal immediate recall; VDR, verbal delayed recall; VIS, visuospatial; EM, episodic memory; HV, hippocampal volume

Table 8. Hierarchical Regression Models with Episodic Memory Subfactors

Model A	β	*p*-value	R^2	Sig.	Model B	β	*p*-value	R^2	Sig.	Model C	β	*p*-value	R^2	Sig.
Total Hippocampal Volume														
1 Age	-.26	<.001**	.199	<.001**										
Gender	<.01	.975												
Race	.07	.078												
Education	-.01	.770												
ICV	.37	<.001**												
2 VIR	.119	.002**	.210	.002**	2 VDR	.122	.002**	.210	.002**	2 VIS	.119	.002**	.210	.002**
3 VDR	.15	.577	.208	.83	3 VIR	-.03	.901	.208	.99	3 VIR	-.03	.901	.208	.79
VIS	<.01	.984			VIS	.15	.577			VDR	.15	.577		
Left Hippocampal Volume														
1 Age	-.27	<.001**	.187	<.001**										
Gender	-.02	.596												
Race	.06	.111												
Education	-.02	.561												
ICV	.34	<.001**												
2 VIR	.13	<.001**	.200	<.001**	2 VDR	.14	<.001**	.201	<.001**	2 VIS	.13	<.001**	.201	<.001**
3 VDR	.31	.258	.200	.42	3 VIR	-.21	.433	.200	.73	3 VIR	-.21	.433	.200	.53
VIS	.04	.850			VIS	.04	.850			VDR	.31	.258		
Right Hippocampal Volume														
1 Age	-.24	<.001**	.183	<.001**										
Gender	.03	.551												
Race	.07	.076												
Education	<.01	.979												
ICV	.37	<.001**												
2 VIR	.099	.012*	.190	.012*	2 VDR	.097	.014*	.190	.014*	2 VIS	.096	.016*	.189	.016*
3 VDR	-.02	.944	.187	.98	3 VIR	.15	.581	.187	.86	3 VIR	.15	.581	.187	.76
VIS	-.03	.880			VIS	-.03	.880			VDR	-.02	.944		

*Statistical significance at the .05 level, **Statistical significance at the .01 level, Notes: Gender = male: 1, female: 2; ICV, intracranial volume; VIR, verbal immediate recall; VDR, verbal delayed recall; VIS, visuospatial. Covariates are included in all steps of all models but are only reported once here to reduce redundancy. R^2 reflects adjusted R^2 values.

Table 9. Hierarchical Regression Models with General Episodic Memory Domain Factor

Model		β	*p*-value	R2	Sig.
Total Hippocampal Volume					
1	Age	-0.26	<.001**	0.199	<.001**
	Gender	<0.01	0.975		
	Race	0.07	0.078		
	Education	-0.01	0.770		
	ICV	0.37	<.001**		
2	EM	0.121	0.002**	0.210	.002**
Left Hippocampal Volume					
1	Age	-0.27	<.001**	0.187	<.001**
	Gender	-0.02	0.596		
	Race	0.06	0.111		
	Education	-0.02	0.561		
	ICV	0.34	<.001**		
2	EM	0.13	<.001**	0.201	<.001**
Right Hippocampal Volume					
1	Age	-0.24	<.001**	0.183	<.001**
	Gender	0.03	0.551		
	Race	0.07	0.076		
	Education	<0.01	0.979		
	ICV	0.37	<.001**		
2	EM	0.098	0.013*	0.190	0.013*

* Statistical significance at the .05 level
** Statistical significance at the .01 level
Notes: Gender = male: 1, female: 2; ICV, intracranial volume; EM, episodic memory. Covariates are included in all steps of all models but are only reported once here to reduce redundancy. R^2 reflects adjusted R^2 values.

Table 10. Hierarchical Regression Models with Hippocampal Subfield Volumes

	Model A	β	*p*-value	R^2	Sig.		Model B	β	*p*-value	R^2	Sig.		Model C	β	*p*-value	R^2	Sig.
CA1 Volume																	
1	Age	-0.28	<.001**	0.171	<.001**												
	Gender	0.05	0.250														
	Race	0.07	0.053														
	Education	<-.01	0.972														
	ICV	0.33	<.001**														
2	VIR	0.115	0.004**	0.181	0.004**	2	VDR	0.116	0.004**	0.181	0.004**	2	VIS	0.114	0.004**	0.180	0.004**
3	VDR	0.09	0.751	0.178	0.94	3	VIR	0.03	0.911	0.178	0.99	3	VIR	0.03	0.911	0.178	0.84
	VIS	<-.01	0.997				VIS	<-.01	0.997				VDR	0.09	0.751		
CA2 Volume																	
1	Age	-0.08	0.042	0.034	<.001**												
	Gender	<0.01	0.924														
	Race	0.02	0.589														
	Education	-0.04	0.337														
	ICV	0.19	<.001**														
2	VIR	-0.04	0.347	0.033	0.35	2	VDR	-0.03	0.470	0.033	0.47	2	VIS	-0.03	0.436	0.033	0.44
3	VDR	0.34	0.250	0.033	0.42	3	VIR	-0.41	0.160	0.033	0.35	3	VIR	-0.41	0.160	0.033	0.37
	VIS	0.03	0.880				VIS	0.03	0.880				VDR	0.34	0.250		
CA3 Volume																	
1	Age	0.01	0.714	0.124	<.001**												
	Gender	-0.03	0.457														
	Race	0.08	0.046*														
	Education	<-.01	0.970														
	ICV	0.32	<.001**														
2	VIR	0.08	0.041*	0.129	0.041*	2	VDR	0.09	0.037*	0.129	0.037*	2	VIS	0.08	0.057	0.128	0.06
3	VDR	0.19	0.513	0.127	0.73	3	VIR	0.03	0.907	0.127	0.80	3	VIR	0.03	0.907	0.127	0.55
	VIS	-0.14	0.513				VIS	-0.14	0.513				VDR	0.19	0.513		
Dentate Gyrus Volume																	
1	Age	-0.20	<.001**	0.096	<.001**												

	Gender	-0.03	0.543													
	Race	-0.02	0.690													
	Education	<0.01	0.931													
	ICV	0.25	<.001**													
2	VIR	0.06	0.140	0.098	0.14	2	VDR	0.07	0.089	0.099	0.09	2	VIS	0.06	0.136	0.098	0.14
3	VDR	0.46	0.113	0.098	0.28	3	VIR	-0.31	0.270	0.098	0.39	3	VIR	-0.31	0.270	0.098	0.28
	VIS	-0.08	0.702				VIS	-0.08	0.702				VDR	0.46	0.113		
Subiculum Volume																	
1	Age	-0.19	<.001**	0.241	<.001**												
	Gender	-0.08	0.069														
	Race	0.14	<.001**														
	Education	-0.06	0.124														
	ICV	0.39	<.001**														
2	VIR	0.154	<.001**	0.260	<.001**	2	VDR	0.145	<.001**	0.257	<.001**	2	VIS	0.155	<.001**	0.260	<.001**
3	VDR	-0.41	0.120	0.260	0.26	3	VIR	0.35	0.172	0.260	0.10	3	VIR	0.35	0.172	0.260	0.28
	VIS	0.21	0.268				VIS	0.21	0.268				VDR	-0.41	0.120		
Entorhinal Cortex Volume																	
1	Age	-0.15	<.001**	0.258	<.001**												
	Gender	-0.15	<.001**														
	Race	0.09	0.016*														
	Education	0.06	0.107														
	ICV	0.37	<.001**														
2	VIR	0.129	<.001**	0.271	<.001**	2	VDR	0.127	<.001**	0.271	<.001**	2	VIS	0.124	.001**	0.270	.001**
3	VDR	0.05	0.846	0.269	0.93	3	VIR	0.15	0.558	0.269	0.83	3	VIR	0.15	0.558	0.269	0.56
	VIS	-0.07	0.704				VIS	-0.07	0.704				VDR	0.05	0.846		

* Statistical significance at the .05 level

** Statistical significance at the .01 level

Notes: Gender = male: 1, female: 2; ICV, intracranial volume; VIR, verbal immediate recall; VDR, verbal delayed recall; VIS, visuospatial. Covariates are included in all steps of all models but are only reported once here to reduce redundancy. R^2 reflects adjusted R^2 values.

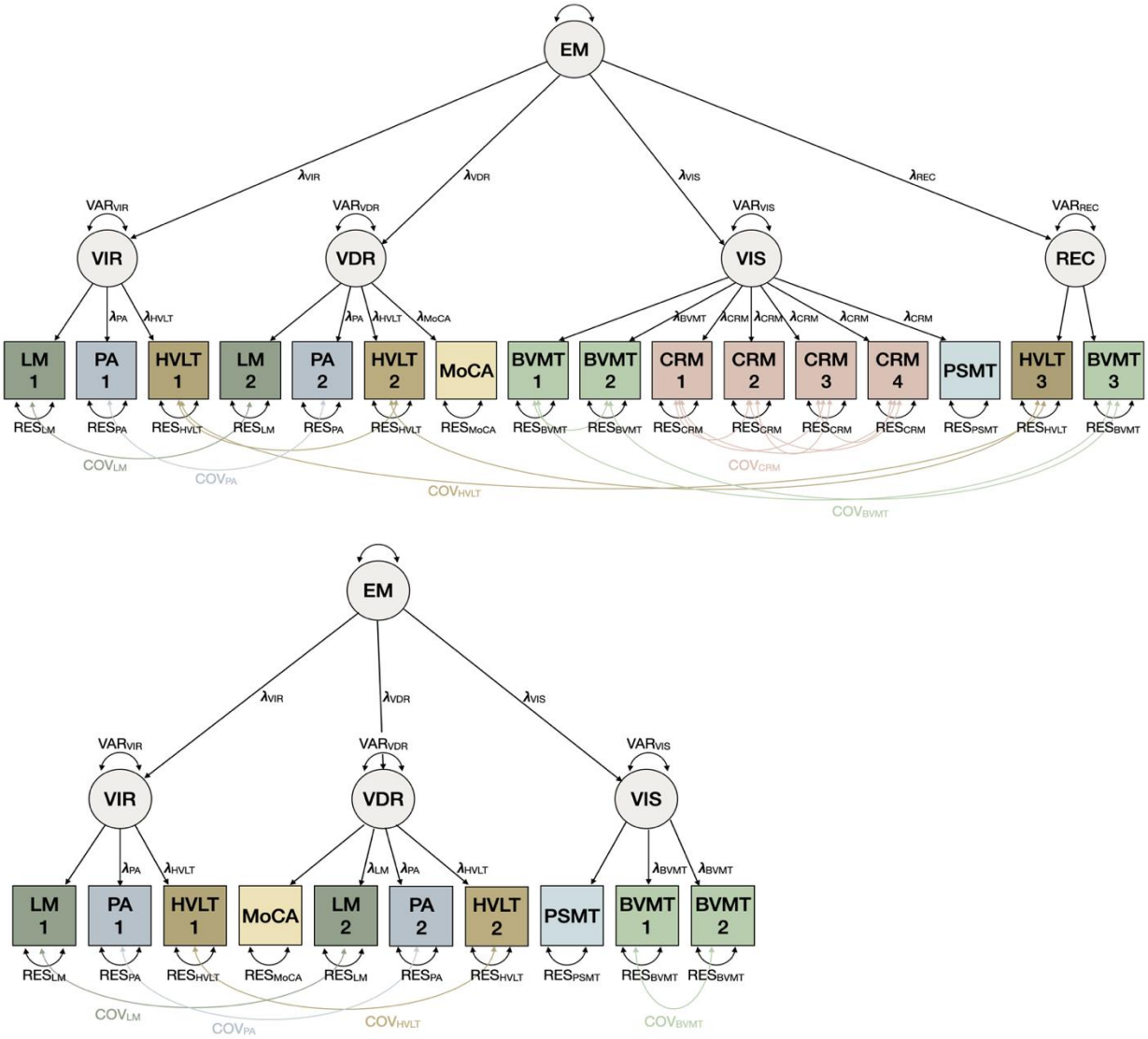

Figure 1. Hypothesized (Top) and Final (Bottom) Confirmatory Factor Analysis Models

EM, episodic memory; VIR, verbal immediate recall; VDR, verbal delayed recall; VIS, visuospatial; REC, recognition; LM, logical memory; PA, paired associates; CRM, Cohen's Relational Memory Test; 1, immediate recall; 2, delayed recall; 3, recognition discrimination index; COV, covariance; VAR, variance; RES, residual; λ, loading. Unlabeled paths are fixed to one.

Appendix B

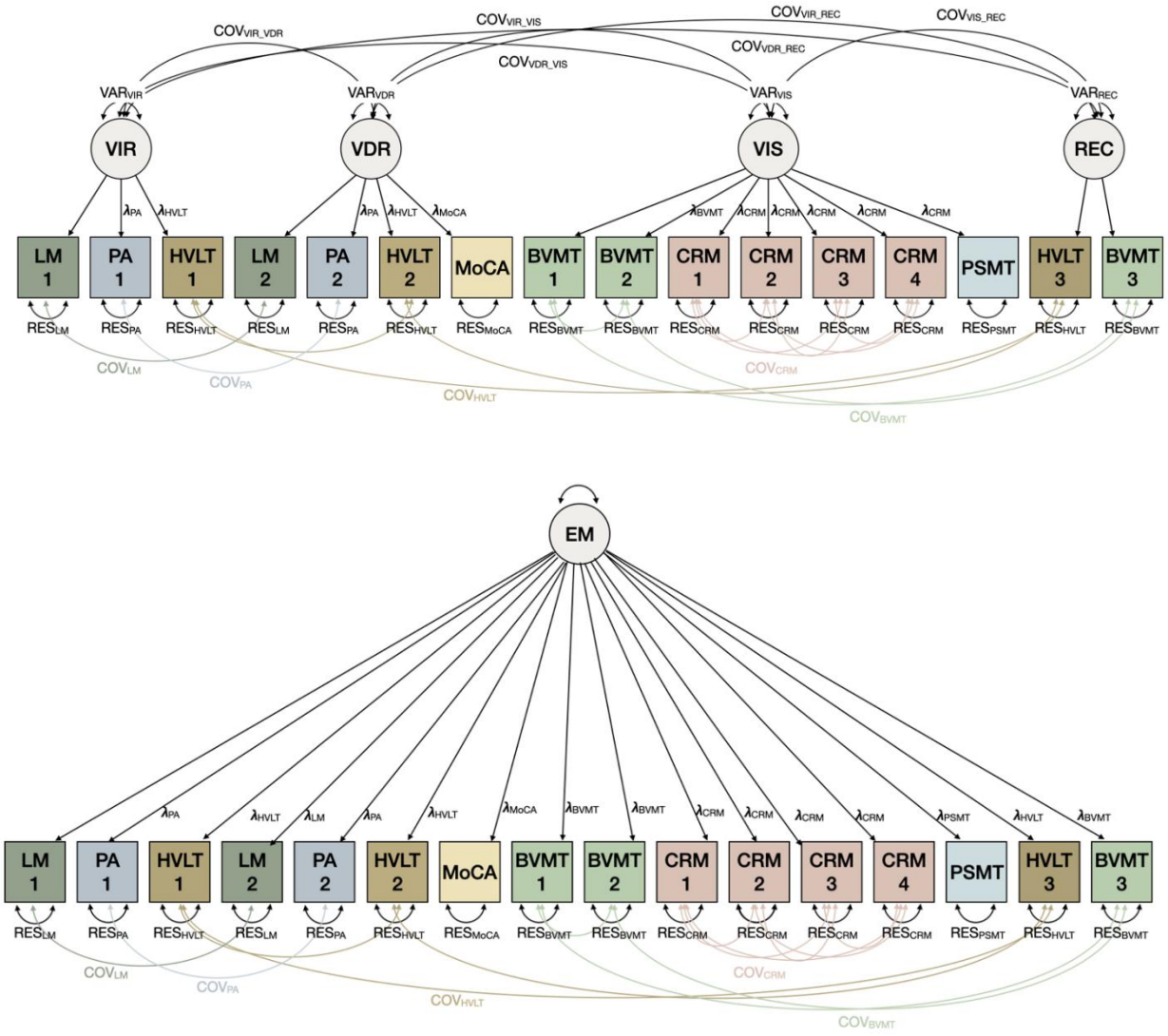

Figure 2. Confirmatory Factor Analysis Models Without General Episodic Memory Factor (Top) and Without Episodic Memory Subfactors (Bottom)

EM, episodic memory; VIR, verbal immediate recall; VDR, verbal delayed recall; VIS, visuospatial; REC, recognition; LM, logical memory; PA, paired associates; CRM, Cohen's Relational Memory Test; 1, immediate recall; 2, delayed recall; 3, recognition discrimination index; COV, covariance; VAR, variance; RES, residual; λ, loading. Unlabeled paths are fixed to one.

Appendix C

Table 11. Factor Loadings Derived from Cohen's Relational Memory Test Confirmatory Factor Analysis Model

Factor	Measure	Loading
Cohen's Relational Memory Test	Misplacement	1.04**
	Edge Resizing	1.07**
	Distortion	0.77**
	Swaps	0.91**

****** Significant at the p<.01 level

Table 12. Hierarchical Regression Models with Cohen's Relational Memory Test Factor

Model		β	*p*-value	R^2	Sig.
Total Hippocampal Volume					
1	Age	-0.26	<.001**	0.199	<.001**
	Gender	<0.01	0.975		
	Race	0.07	0.078		
	Education	-0.01	0.770		
	ICV	0.37	<.001**		
2	CRMT	0.005	0.899	0.198	0.899
Left Hippocampal Volume					
1	Age	-0.27	<.001**	0.187	<.001**
	Gender	-0.02	0.596		
	Race	0.06	0.111		
	Education	-0.02	0.561		
	ICV	0.34	<.001**		
2	CRMT	-0.003	0.944	0.186	0.944
Right Hippocampal Volume					
1	Age	-0.24	<.001**	0.183	<.001**
	Gender	0.03	0.551		
	Race	0.07	0.076		
	Education	<0.01	0.979		
	ICV	0.37	<.001**		
2	CRMT	0.01	0.752	0.182	0.752

* Statistical significance at the .05 level

** Statistical significance at the .01 level

Notes: Gender = male: 1, female: 2; ICV, intracranial volume; CRMT, Cohen's Relational Memory Test. Covariates are included in all steps of all models but are only reported once here to reduce redundancy. R^2 reflects adjusted R^2 values.