



Published in final edited form as:

*Clin Neuropsychol.* 2020 January ; 34(1): 158–173. doi:10.1080/13854046.2018.1538428.

## A Tale of Two Stories: Validity of an Alternative Story Memory Test in a Sample of Older Adults

Erin Trifilio<sup>1</sup>, Jared J. Tanner<sup>1</sup>, London Butterfield<sup>2</sup>, Paul Mangal<sup>1</sup>, Jacqueline E. Maye<sup>1</sup>, Michael Marsiske<sup>1</sup>, Catherine C. Price<sup>1</sup>, Dawn Bowers<sup>1,3</sup>

<sup>1</sup>Clinical and Health Psychology, University of Florida, Gainesville FL, USA

<sup>2</sup>BayCare Medical Group, Suncoast Medical Clinic, Saint Petersburg FL, USA

<sup>3</sup>Neurology, University of Florida

### Abstract

**Objective:** Many patients require repeat neuropsychological evaluations to determine change over time. Repeat evaluations lead to practice effects, which can impact the validity of the assessment. The current study assessed, in older adults, the validity of an alternative set of verbal memory stories created by Newcomer and colleagues (1994).

**Method:** A total 154 of nondemented adults, ages 60 to 92, completed the WMS-III Logical Memory (LM) stories and two Newcomer stories (Carson-Jones) as part of a larger battery of neurocognitive tests. The Carson-Jones stories were scored for: 1) verbatim (traditional) and 2) thematic (developed for this study) accuracy. Story memory variables were compared to each other and additional neurocognitive measures using bivariate correlations. A subset of participants ( $n=133$ ) completed magnetic resonance imaging (MRI) and various structural regions (e.g., thickness and volume of medial temporal lobe structures) were used to assess external validity of Carson-Jones stories with hierarchical multiple regression analyses.

**Results:** There was a strong positive correlation between WMS-III LM and Carson-Jones stories for both verbatim and thematic scoring. Both scoring types showed convergent validity with other verbal memory measures (e.g., WMS-III LM and HVLT-R Delay/Learning) and divergent validity with Stroop Word Reading and JOLO. Regarding neuroimaging correlates, Carson-Jones verbatim scoring was significantly associated with left subiculum and left whole hippocampal volume whereas thematic scoring was significantly associated only with left subiculum.

**Conclusions:** Newcomer stories appear to be a valid alternative to WMS-III LM stories in terms of assessing verbal memory in healthy older adults.

### Keywords

memory; older adults; mental status and dementia tests; hippocampus; validation

## Introduction

Neuropsychologists frequently use story memory tests to assess verbal memory. Story memory tests were designed to assess the integrity of the medial temporal lobe memory system and have been validated in this regard (Wechsler, 1945; Squire, 1987). Traditionally, examinees are read stories and then asked to recall them immediately and then later, after a 20–30 minute delay. The most widely known and used of these measures are the Logical Memory stories from the Wechsler Memory Scale (WMS). The original version of the WMS (1945) included two alternate sets of stories, which were sometimes interchanged to minimize practice effects. However, subsequent versions of the WMS (R, III, IV) reverted to only one set of stories, which posed difficulty when patients were tested on multiple occasions.

A vast literature over the past 20 years shows substantial practice effects for Logical Memory stories given over multiple visits to healthy younger and older adults and patients. Practice effect magnitudes (e.g.,  $\beta$ , Cohen's  $d$ ) for both WMS-III and WMS-IV versions have ranged from moderate to large across studies (Theisen, Rapport, Axelrod, & Brines, 1998; McCaffrey, Ortega, Orsillo, Nelles, & Haase, 1992; Gavett, Ashendorf, & Gurnani, 2015; Gavett et al., 2016). Factors influencing these effect sizes include the time between test sessions (i.e., 2 weeks, etc.), number of exposures to test stimuli (i.e., prior testing instances), the types of patients evaluated (e.g., normal, MCI, demented), and characteristics of the test itself (Salthouse, Schroeder, & Ferrer, 2004; McCaffrey et al., 2000; Lezak, Howieson, & Loring, 2004). For example, practice effects on verbal memory tests (e.g., story memory, list learning) are less pronounced amongst individuals with Alzheimer disease than in MCI and healthy adults (Cooper et al., 2001; Duff et al., 2007).

In neuropsychology, one critical goal is to determine whether meaningful cognitive changes are occurring over time. Assessing cognitive change requires repeated testing, often with the same or alternative forms of the same measure. A reliable change index (RCI) is one way to compute change and refers to the amount of change one would expect in a test score due solely to repeat evaluation (Jacobson & Truax, 1991; Chelune, Naugle, Luders, Sedlak, & Awad, 1993; Iverson et al., 2001). The RCI calculations vary in complexity but take into account baseline scores, standard error, and change in scores over time in order to produce a value indicative of change unrelated to chance fluctuation. If changes in test performance are large enough (i.e., greater than the reliable change cutoff of 90% or 95% confidence interval), then they can be considered statistically significant. Though reliable change indices can improve clinical interpretation of change over time by addressing bias, reliability, and sometimes group practice effects, they do not fully address the influence of repeat evaluations using the same test stimuli on an individual level. Ideally, patients returning for a repeat evaluation would be administered an alternate story set in lieu of the LM stories in order to avoid overt practice effects, which could hinder interpretation of the results and cloud clinical judgment.

Several alternate story paragraph tests have been developed for use in place of traditional WMS LM paragraphs for the purposes of repeat testing (Schnabel, 2012; Sullivan, 2005; Morris et al., 1997; Newcomer, Craft, Hershey, Askins, & Bardgett, 1994). However, most

have been developed and normed primarily in young, generally healthy adults, particularly college students and middle-aged controls. By contrast and of particular relevance to the current study, Newcomer and colleagues developed a set of eight brief paragraphs for use in research studies of stress-induced glucocorticoid-related cognitive impairment and schizophrenia (Newcomer et al., 1994, 1999a, 1999b). No ‘classic’ normative data were provided, though information about the performance of the relatively small control groups ( $n = 9, 14, & 20$  across studies) were given in terms of means and standard deviations for immediate and delayed recall. Of note, these prior studies using the Newcomer story series used the original ‘verbatim’ scoring method, with points awarded for each ‘bit’ remembered (44 bits total; Newcomer et al., 1994). This differs from the LM scoring method in that LM relies more heavily upon thematic recall and less on specific word/phrase recall. The authors did develop an alternative to the verbatim scoring method called ‘paraphrase’ scoring. Even so, this method still relies on scoring each individual word rather than ‘gist’ phrases. Scoring wise, to make the two story sets more comparable would require development of ‘gist’ or thematic criteria for the Newcomer stories that more closely mirrors the LM criteria.

The goal of the present study was to learn whether a subset of two stories, taken from the Newcomer series, would serve as a valid alternative to Logical Memory Stories for clinical assessment of recent memory in older adults. There were three specific aims:

*Aim 1* addressed the scoring method discrepancies between the two story sets. We wanted to learn whether a newly developed thematic scoring method of NS correlates with LM better than the verbatim system developed by Newcomer et al (1994). We predicted that the NS thematic scoring would correlate with LM traditional scoring as well as or better than verbatim traditional NS scoring. As part of this aim, we also assessed interrater reliability and internal consistency using the new thematic scoring type.

*Aim 2* examined the convergent and divergent validity of a subset of the Newcomer stories. This was done by comparing performance on WMS-III Logical Memory stories to a subset of the Newcomer stories. Due to pragmatic constraints, we selected two of the eight Newcomer stories, the Lucy Carson and the Adam Jones stories, for the present study. If the Newcomer stories are, indeed, a valid verbal memory test assessing medial temporal lobe functioning similarly to WMS-III LM, then the two story sets should be highly correlated with each other, thus providing evidence for convergent validity. In terms of divergent validity, based on prior factor and correlational analyses identifying discrete domains separating memory from other cognitive domains, (Larrabee, Kane, & Shuck, 1983; Leonberger, Nicks, Larrabee, & Goldfader, 1992; Tulskey & Price, 2003), we predicted that performance on Newcomer stories would correlate less well with tests of visuospatial ability and speeded word reading than verbal memory tests.

*Aim 3* examined the external validity of the Newcomer stories (NS) by evaluating the relationship between delayed memory recall scores and volume of mesial temporal regions (versus control regions). This was done with a subset of participants ( $n = 133$ ) who also completed structural neuroimaging. We predicted that the NS performance would correlate better with left hemisphere medial temporal lobe structures, which are crucial to verbal memory formation and recall (e.g., entorhinal cortex, whole hippocampus, subiculum) and

would be less strongly correlated with other ‘non-memory’ regions (e.g., rostral middle frontal) within the left hemisphere. We also predicted that the NS would correlate more strongly with left than right temporal lobe regions due to hemispheric laterality effects in processing verbal versus nonverbal information.

## Methods

### Participants

Participants included 192 older adults between the ages of 60 and 92 years who were healthy participants from two ongoing IRB approved studies at the University of Florida. One study (Bowers, Marsiske – ReVitalize) included 52 healthy older adults enrolled in a cognitive intervention (though no intervention effects were examined for the current study) program (Sample 1) and the other study (Price) included 140 non-demented (of which, 21 were identified post-hoc as meeting criteria for MCI) older adults enrolled in prospective serial cognitive assessment with neuroimaging (Sample 2). Inclusion criteria for the current analyses: 1) 60 years of age, 2) able to read and write at an 8<sup>th</sup> grade level, and 3) intact instrumental activities of daily living (Lawton & Brody, 1969). Exclusion criteria: 1) previous major strokes or other known significant brain abnormality, 2) evidence of potential dementia (e.g., Mini-Mental Status Examination scores < 24 or Montreal Cognitive Assessment scores < 20), 3) current or past history of major psychiatric disturbance or alcohol/substance abuse within the past six months, and 4) significant medical disease (e.g., cancer requiring treatment within the last 5 years, chronic hepatitis, history of organ transplant, etc.).

### Procedures

As part of their parent studies, all participants received a neuropsychological battery that included a cognitive screening measure, either the Mini Mental Status Exam (MMSE; Folstein, Folstein, & McHugh, 1975; Sample 1) or Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005; Sample 2), two stories from the Newcomer series (described below), and measures of: a) *recent memory* including, WMS-III Logical Memory Stories (WMS-III LM; Wechsler, 1945) and Hopkins Verbal learning Test-Revised (HVLT-R; Brandt & Benedict, 1991), b) *visuospatial skills*, Judgement of Line Orientation (JOLO; Benton, Hamsher, Varney, & Spreen, 1983), c) *executive function*, the Stroop (Golden & Freshwater, 2002); d) and a measure that estimated IQ based on single word reading, either the Wechsler Test of Adult Reading (WTAR; Wechsler, 2001) or the Word Reading Subtest of the Wide Range Achievement Test 4 (WRAT 4; Wilkinson & Robertson, 2006). Each participant also completed standardized self-report mood measures assessing a) *depression* including the Beck Depression Inventory, 2<sup>nd</sup> Edition (BDI-II; Beck, Steer, & Brown, 1996) and the Geriatric Depression Scale (GDS; Yesavage et al., 1982); b) *anxiety* including State Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983); and c) *apathy* including Apathy Scale (AS; Starkstein et al., 1992).

**Newcomer Story Set**—For the present study, participants received two of the Newcomer stories, referred to as Carson-Jones (See Figures 1 and 2). These particular stories were chosen, as they are the first two in the sequence of eight stories, the ‘Lucy Carson’ story and

the “Adam Jones” story. Information about required reading level and general complexity is as follows: *Lucy Carson Story*: Flesch-Kincaid grade level = 7.4, 61 total words, 4 total sentences, 15.25 words per sentence, 1.44 syllabus per word, and 44 total ‘bits’; *Adam Jones’ story*: Flesch-Kincaid grade level = 6.4, 62 total words, 4 total sentences, 15.5 words per sentence, 1.35 syllables per word, and 44 total ‘bits’). For comparison, the WMS-III LM stories are of similar complexity in terms of grade level but with two more sentences and more total words in the Joe Garcia story. *Anna Thompson Story*: Flesch-Kincaid grade level = 8.1, 65 total words, 3 total sentences, 21.67 words per sentence, 1.29 syllables per word, and 25 scoring ‘bits’; *Joe Garcia Story*: Flesch-Kincaid grade level = 7.7, 86 total words, 6 total sentences, 17.2 words per sentence, 1.41 syllables per word, and 25 scoring ‘bits’.

Each story was read aloud by a trained examiner followed by immediate and 30 minute delay free recall conditions. As with the WMS-III LM stories, participants were cued to remember the Carson-Jones stories for subsequent delayed recall. The stories were recorded verbatim for later scoring. Scoring of the Newcomer stories was done in two ways – verbatim according to the procedures developed by Newcomer and thematic scoring, using guidelines developed for this study.

**Verbatim Scoring.:** One full point (1.0) for each bit was awarded for each ‘perfect verbatim response.’ In this situation, the content words for the bit had to be recalled exactly and completely. Non-content words (i.e., articles such as, ‘and’, ‘the’, and, ‘a’) were not required for full credit. Partial credit (0.5 points) was given for recall of a word with the same lexical root and phoneme (i.e., ‘car’ for ‘cars’). The order of recall did not factor into scoring (i.e., ‘plants’ could be recalled in the original sentence or another with full credit awarded). Points were not awarded for any other recall type. For example, no points were given for verbs recalled in the wrong tense if the root word changed (i.e., ‘go’ for ‘went’), nouns recalled as pronouns (i.e., ‘her’ for ‘Lucy’), or incorrect name or number (i.e., ‘25’ for ‘35’).

**Thematic Scoring.:** A thematic-type scoring system was developed that focused on synonyms or ‘gist’ recall rather than recall of identical words. For example, scoring criteria for the phrase, ‘got lost,’ is ‘indication that the main character lost their way.’ This was developed in order to more accurately compare the Newcomer and WMS story scores. The total number of thematic units was 28 for the Lucy Carson story and 26 for Adam Jones.

**Order of Administration.:** Neurocognitive measures and other measures including mood measures were given over multiple days; story memory tasks (WMS Logical Memory and Carson-Jones) were given on different days, with WMS stories always given first.

**MR Procedures—**A subset of 133 participants completed a brain MRI within a 3T Siemens Verio (8 channel head coil) with isovoxel T1-Weighted sequences: TR: 2500ms; TE: 3.77ms; 176 sagittal 1mm<sup>3</sup> slices, 1 mm isotropic resolution; 256×256×176 matrix, 7/8 phase partial Fourier, total acquisition time: 9:22. T1-weighted images were post-processed using an automated longitudinal pipeline (FreeSurfer 6.0; Fischl, 2012; Fischl et al., 2002). Pial surfaces, gray and white matter boundaries and structure segmentations were checked for reconstruction accuracy. Cortical reconstructions were automatically parcellated using the Desikan-Killiany-Tourville atlas (Fischl et al., 2004; Desikan et al., 2006; Klein &

Tourville, 2012) Hippocampal subfields were also extracted using FreeSurfer (-hippocampal-subfields-T1).

**Post-hoc Identification of Mild Cognitive Impairment**—In order to increase variability among neuroimaging measures, we included data from 21 older adults who had been excluded from the normative aims because they performed below age/education-adjusted normative values for our cognitive tests. These 21 participants met criteria for a ‘diagnosis’ of mild cognitive impairment (MCI) using the comprehensive criteria outlined by Jak and colleagues (2009). Participants were classified as MCI if they scored 1 standard deviation below age-appropriate normative data on two or more measures within a cognitive domain. In order to calculate normative data for only those who performed normally on our tests, analyses were run excluding these individuals for Aims 1 and 2 but included for Aim 3 to assess a wider range of brain-behavior relationships.

### Statistical Analyses

The two stories in the Carson-Jones set were combined to create separate composite scores (i.e., sum of the two stories) for immediate and delay recall conditions. This was done for both the verbatim and thematic scoring systems in light of the high inter-story correlation within the verbatim ( $r = 0.684$  Immediate,  $0.720$  Delay) and thematic ( $r = 0.623$  Immediate,  $r = 0.660$  Delay) conditions. Similarly, both stories of WMS-III LM set were combined to create an immediate recall composite and a delayed recall composite, again due to their strong inter-story correlation ( $r = 0.528$  Immediate,  $r = 0.593$  Delay) conditions. These composites were used for all subsequent analyses. For the newly-developed thematic scoring, we assessed internal consistency using Cronbach’s alpha and interrater reliability using intraclass correlation analysis (threshold at  $0.80$ ). To compare the magnitude of the correlations between WMS-III LM and Carson-Jones scoring types (e.g., verbatim and thematic) we used Steiger’s Z test (which enables comparison of dependent correlations). We examined convergent and divergent validity of the Carson-Jones stories using Pearson bivariate correlations between Carson-Jones, WMS-III LM, and additional neurocognitive tests in memory and non-memory domains. To be conservative, we selected a threshold of 25% shared variance (i.e.,  $r=0.5$ ) was established to define “meaningful” associations between measures. Normative values (e.g., T scores and percentiles) for Carson-Jones stories thematic scoring (excluding MCI participants) were calculated for age- and education-specific groups (See Supplemental Material for norms tables).

Finally, we examined the relationship between delayed recall and structural brain variables in a subset of participants with neuroimaging data (133 had complete neuroimaging data). Only delayed recall measures were used in these analyses, due to our interest in ‘retention’ of information over time. These data were analyzed using hierarchical multiple regression with pair-wise deletion (all 133 participants were included for these analyses), controlling for age and years of formal education. The following brain regions were chosen due to their role in verbal memory and their sensitivity to age-related changes: left entorhinal cortex thickness, left whole hippocampal volume, and left subiculum volume. All volume measurements were corrected for total intracranial volume (ICV) by dividing each measurement by the participant’s total ICV. Corresponding right temporal regions were



selected and predicted to be less related to story memory performance as the left hemisphere primarily mediates verbal functions, even in the majority of left-handed individuals. Thickness of the rostral middle frontal cortex (left and right hemispheres) was chosen as a control region as it is thought to be less related to memory processes and more so to executive functioning (i.e., working memory). Individual multiple linear regressions were run for each predicted brain structure and story memory test with the brain region of interest as the DV and story memory task as IV (Block 2), controlling for age and education in Block 1. This analysis served as a means of assessing external validity of the Carson-Jones stories. We compared the magnitudes of the relationships between the two story memory tests and structural brain regions involved in memory and retrieval. We did this by observing the  $R^2$  incremental change when adding each story memory variable to a baseline model including age and education as control variables.

## Results

### Participants

Sample characteristics are shown in Tables 1 and 2. Overall, 36 had missing story data. Those with missing data were comparable in terms of age ( $p = .367$ ), education ( $p = .570$ ), and gender ( $p = .700$ ) to those with complete data. These individuals were removed along with two participants who had BDI-II/GDS scores  $> 19$  (above mild depression threshold), one individual who had a MoCA score  $< 20$ , and those who met criteria for mild cognitive impairment using comprehensive criteria (Jak et al., 2009) resulting in a total of 132 participants. Figure 3 depicts the sample sizes and data exclusion for each aim. Overall, the participants were well educated, ranged in age from 60 to 92, and consisted of 39% Males and 61% Females. We examined whether the participants drawn from the two separate funded investigations differed in terms of demographic variables. As shown in Table 1, we found that the two sample groups (1 & 2) were similar except for slight differences in age and education, with Sample 1 (intervention group) being minimally older with more years of education. Even so, both groups represented a wide age range and were generally highly educated.

Medication prescriptions had the following percent frequency: antianxiety (12%), antidepressant (17%), and anticholinergics (33%). Average anticholinergic burden (ACB) score using the Magellan Anticholinergic Risk Scale was 0.768 (1=low risk, 2=moderate risk, 3=high risk of anticholinergic side effects) and only 13% of participants had a total ACB score of  $\geq 3$ . On a health screening measure, the Charlson Comorbidity Index (CCI), the overall mean score was less than 1, indicating that the majority of individuals scored either a 0 or a 1, indicating good overall health and lack of major medical problems. Individual self-report medical history was also reviewed.

### Aim 1: New Thematic Scoring

For Aim 1, we assessed the relationship between newly developed thematic scoring criteria and existing verbatim criteria and WMS-III LM stories within only those participants identified as cognitively healthy. The intraclass correlation coefficient and Cronbach's alpha between two independent scorers for the new thematic scoring of the Carson-Jones was

appropriately high for both stories (Lucy Carson & Adam Jones) for both immediate (ICC = 0.981–0.983; alpha = 0.988–0.990) and delayed (ICC = 0.969–0.983; alpha = 0.980–0.990) recall conditions. Steiger's Z tests were non-significant when comparing the strength of correlation of WMS-III LM scores to both verbatim and thematic scoring of Carson-Jones stories (Immediate  $z = .645, p = .519$ ; Delay  $z = .098, p = .922$ ).

## **Aim 2: Convergent and Divergent Validity**

**Convergent Validity – Story Memory Tests and Other Memory (HVL-R)**—For Aim 2 we assessed the relationship between Carson-Jones stories and other verbal memory measures to determine convergent validity within the cognitively healthy participants. Pearson bivariate correlations between the two story tests (e.g., WMS-III LM & Carson-Jones stories) were significant ( $p < .01$ ) for both immediate and delay recall conditions and for both scoring types for Carson-Jones (see Table 3). In general, correlations were highest between immediate and delay conditions within story types (e.g., between LM I & II and between Carson-Jones Imm. & Delay). However, there was also a strong relationship between the two scoring types for Carson-Jones (e.g., between Carson-Jones Imm-V & Carson-Jones Imm-T). In addition, story memory tests correlated significantly with another verbal memory measure (HVL-R) in both delayed and learning conditions.

**Divergent Validity – Other Cognitive Domains**—Aim 2 also assessed the relationship between Carson-Jones stories and other 'non-memory' measures to determine divergent validity. Pearson bivariate correlations were calculated for combined immediate and delay scores for each story type. Logical Memory stories showed significant but weak associations with non-memory domain tests (e.g., Stroop Word Reading, JOLO). Carson-Jones stories were not significantly associated with these measures, providing evidence for divergent validity. Results for convergent and divergent validity correlations were examined using both simple bivariate correlations and as partial correlations while controlling for total ACB. The patterns of significance and effect sizes did not differ appreciably. Results presented here are simple correlations without controlling for effects of anticholinergic medication.

## **Aim 3: Relationship with Structural Brain Variables**

We examined external validity by assessing the relationship between story memory tests and structural MR variables of both memory and non-memory regions. We also wanted to assess whether Carson-Jones stories were more strongly associated with memory regions than WMS-III. For these analyses, both groups of participants – healthy and those meeting criteria for MCI – were included. Across all regression equations, education and age accounted for a large amount of variance in structure-story recall relationships. Even so, we found that verbatim scoring of Carson-Jones (Delay) added a significant amount of variance to the model predicting both left subiculum and left whole hippocampal volume. Logical Memory Delay also added significant variance to the models predicting left subiculum volume as well as whole hippocampal volume. Carson-Jones Delay thematic scoring significantly predicted left subiculum volume only. Though the effect sizes for Carson-Jones stories were small across regression analyses, this was true for WMS-III LM stories as well, for which we found only two significant effects. As predicted, none of the story memory tests significantly predicted left rostral middle frontal thickness. Most importantly, we did



not find relationships between right hemisphere structures and story variables with the exception of trend level significance with LM Delay predicting right subiculum volume. In the same analyses using standardized z rather than raw scores, comparison of the confidence intervals did not reveal significant differences between the b-weights for any of the aforementioned relationships, suggesting associations of similar magnitudes across story and scoring types.

## Discussion

The overall aim was to assess the use of a new set of story paragraphs, the Carson-Jones stories, for assessing verbal memory in a sample of healthy older adults. This sample was very healthy and on few medications, though relatively homogenous in terms of education and race. There were three major findings.

First, Carson-Jones stories were scored using two scoring variations: established verbatim scoring criteria and a newly-developed thematic criteria. Verbatim scoring was highly correlated with thematic scoring and both scoring protocols showed similar magnitude associations with LM. As the thematic scoring criteria more closely mirrors LM scoring, this was somewhat unexpected. Even so, these results follow our prediction that the thematic scoring criteria for Carson-Jones should be correlated with the LM stories at least as well as the verbatim scoring.

Second, we found evidence for convergent and divergent validity of the Carson-Jones stories. In terms of convergent validity, Carson-Jones stories showed moderate to strong associations with other verbal memory measures, including WMS-III LM. However, LM and Carson-Jones delayed recall scores were less strongly correlated than immediate. Evidence for divergent validity was supported by absent correlations with tests in other cognitive domains for (e.g., JOLO and Stroop Word Reading) for the Carson-Jones stories.

The Carson-Jones stories were moderately correlated with LM stories, suggesting their utility in clinical repeat testing. However, it is worth considering why these paragraph memory tests are not more highly correlated (i.e.,  $> 0.8$ ). One possibility is that the two story tests vary in their overall complexity, which would result in slightly disparate scores despite their similar underlying neural processes. Though, the stories are comparable in terms of most of their complexity indices (e.g., Flesch Kincaid Grade Level, number of words and sentences), the Joe Garcia story from LM has two more sentences and more words overall than both of the Carson-Jones stories. Another explanation is that these paragraphs vary in their emotional content (i.e., a robbery of a woman's rent money vs. a story about getting lost in a new city). It is well established that stimuli can be made more or less memorable depending upon its emotionality (Eysenck, 1976; Heuer & Reisberg, 1990; Christianson & Loftus, 1991). Finally, there are administration differences, in that the Carson-Jones stories are administered only one time during the learning phase and thus delayed recall does not benefit from the added repetition afforded the LM stories.

Third, we found support for a significant positive relationship between delayed recall of the Carson-Jones stories (especially verbatim recall) and overall volume of the left mesial

temporal region and the left subiculum, in particular. The subiculum is one of the major output structures of the hippocampus from which the fornix arises within the medial temporal lobe system. The memory performance-left mesial temporal relationship was statistically significant, albeit it small. Unlike Price et al. (2010), we did not find a relationship with the left entorhinal region, though this may be due to methodological differences (e.g., combining immediate and delayed recall, etc.). Importantly, delayed memory recall in the current study was not associated with volume or thickness of right mesial temporal lobe regions, nor with ‘nonmemory’ regions in the frontal lobe (i.e., rostral midfrontal). A similar relationship with the left subiculum was found for delayed recall of the WMS-III stories, which also correlated with left whole hippocampal volume. A number of factors may influence the relationship between medial temporal lobe structures and verbal memory performance. For example, the subiculum has been found to be sensitive to ‘healthy’ age-related changes (Chetlat et al., 2008) and also a possible marker of early Alzheimer disease pathology (Carlesimo et al., 2015). This suggests that there are specific age-related as well as neurodegenerative changes that may be affecting these relationships. It is possible that these ‘healthy’ individuals may go on to later develop cognitive impairment and may be experiencing early changes in brain structures that have not yet caused cognitive changes. Alternatively, differences in subiculum and presubiculum volumes may represent a spectrum of age-related changes, ultimately reflecting differential memory performance.

The participants in the current study were healthy, educated, and primarily Caucasian older adults ranging in age from 60 to 92 years. Even though results of this study are likely not confounded by health factors, the applicability of these normative data is limited given the education, race and ethnicity of the participants. For example, normative data tables have exceedingly small sample sizes for certain age-education groups (e.g., 80–89 years old with 12–15 years of education). An additional limitation relates to discrepancies in administration instructions between WMS-III LM and Carson-Jones stories. The instructions for WMS-III LM administration for older adults include two immediate recall conditions (i.e., two readings) of the Story 2 (Joe Garcia), whereas the Carson-Jones stories include just one immediate recall trial for each story, conferring a benefit of second repetition when attempting to recall Story 2 of WMS-III LM. Findings are also limited by order of administration bias (i.e., LM stories were always administered first). Finally, future studies with this type of data should calculate test-retest reliability of Carson-Jones stories as well as intrarater scoring reliability.

Even so, to our knowledge, this study is the first to utilize a large sample of healthy older adults with multiple cognitive measures, including a widely used story memory test (LM) and structural imaging data in order to validate the Newcomer paragraph story tests. With access to alternative tests to measure declarative episodic verbal memory, clinicians and researchers will have the ability to more accurately measure change over time in memory-related processes.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Appendix

**Appendix Table 1.**

Normative values (T-score and %ile) for raw scores for Carson-Jones Thematic Scoring [Age range 60–69; Education 12–15 years].

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
12.5	5	33	7.5	5	33
15	9	37	9.5	9	37
17.5	14	39	10	14	39
21.5	18	41	12	18	41
23	23	43	20	23	43
23.5	27	44	20.5	27	44
25.5	32	45	21	41	47
26.5	36	46	22	46	49
27	41	47	23.5	50	50
27.5	46	49	24	59	52
28	50	50	25	64	53
29.5	55	51	29	68	55
30	59	52	29.5	73	56
31	64	53	32	82	59
32	68	55	32.5	91	63
33	73	56	42	99	72
36	77	57			
37	82	59			
40	91	63			
46	96	67			
48	99	72			

N = 22

**Appendix Table 2.**

Normative values (T-score and %ile) for raw scores for Carson-Jones Thematic Scoring [Age range 70–79; Education 12–15 years].

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
14.5	7	35	10	7	35
18	13	39	15	13	39
27	27	44	17.5	20	41
29	40	47	19.5	27	44
30	47	49	21	33	45

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
32.5	53	51	24	40	47
33	60	52	27	53	51
34.5	67	54	27.5	60	52
35.5	73	56	28.5	67	54
36	80	58	31	80	58
36.5	87	61	33.5	87	61
38	93	65	37.5	93	65
39.5	99	72	39	99	72

N = 15

**Appendix Table 3.**

Normative values (T-score and %ile) for raw scores for Carson-Jones Thematic Scoring [Age range 80–89; Education 12–15 years].

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
23.5	33	45	12.5	67	54
29	67	54	23.5	99	72
30	99	72			

N = 3

**Appendix Table 4.**

Normative values (T-score and %ile) for raw scores for Carson-Jones Thematic Scoring [Age range 60–69; Education 16 years].

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
Raw	%ile	T-score	raw	%ile	T-score
21.5	7	35	19	7	35
25.5	14	39	21	14	39
26	21	42	22	29	44
26.5	29	44	22.5	36	46
27	36	46	24	43	48
28	43	48	27	50	50
28.5	50	50	28	64	53
32	57	51	33.5	71	55
36.5	64	53	36	86	61

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
Raw	%ile	T-score	raw	%ile	T-score
38	71	55	36.5	93	65
40	79	58	44.5	99	72
42	93	65			
46.5	99	72			

N = 14

**Appendix Table 5.**

Normative values (T-score and %ile) for raw scores for Carson-Jones Thematic Scoring [Age range 70–79; Education 16 years].

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
11.5	8	36	8	8	36
25	17	40	21	17	40
26.5	25	43	22	25	43
27.5	33	45	23	33	45
30	42	48	24	42	48
32	50	50	26.5	50	50
35	58	52	31	58	52
36	67	54	31.5	67	54
37	75	57	33.5	83	59
38.5	83	59	36	92	64
39	92	64	36.5	99	72
44	99	72			

N = 12

**Appendix Table 6.**

Normative values (T-score and %ile) for raw scores for Carson-Jones Thematic Scoring [Age range 80–89; Education 16 years].

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
10	14	39	7	14	39
16	29	44	9	29	44
19.5	43	48	14	43	48
21	57	51	17.5	57	51

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
<b>24</b>	71	55	<b>19</b>	71	55
<b>29</b>	86	61	<b>24</b>	86	61
<b>43</b>	99	72	<b>41</b>	99	72

N = 7

**Appendix Table 7.**

Normative values (T-score and %ile) for raw scores for Carson-Jones Thematic Scoring [Age range 60–69; Education 17–20 years].

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
<b>19.5</b>	5	33	<b>13</b>	5	33
<b>27</b>	24	43	<b>17.5</b>	10	37
<b>27.5</b>	33	45	<b>19</b>	14	39
<b>28</b>	38	47	<b>19.5</b>	19	41
<b>28.5</b>	43	48	<b>21.5</b>	24	43
<b>29</b>	52	51	<b>22.5</b>	29	44
<b>30</b>	57	51	<b>23</b>	33	45
<b>31</b>	62	53	<b>24</b>	38	47
<b>31.5</b>	67	54	<b>24.5</b>	48	49
<b>32</b>	71	55	<b>25.5</b>	52	50
<b>33</b>	76	57	<b>26.5</b>	57	51
<b>33.5</b>	81	59	<b>28</b>	62	53
<b>36</b>	86	61	<b>28.5</b>	67	54
<b>37</b>	91	63	<b>29</b>	71	55
<b>42.5</b>	95	66	<b>29.5</b>	76	57
<b>44</b>	99	72	<b>32</b>	81	59
			<b>32.5</b>	86	61
			<b>36.5</b>	91	63
			<b>40</b>	95	66
			<b>41</b>	99	72

N = 21



**Appendix Table 8.**

Normative values (T-score and %ile) for raw scores for Carson-Jones Thematic Scoring [Age range 70–79; Education 17–20 years].

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
13.5	4	32	8	4	32
15.5	8	36	11	8	36
24	15	39	16	15	39
24.5	19	41	17.5	23	43
25	23	43	18	27	44
25.5	31	45	20	35	46
26.5	39	47	21	39	47
28	54	51	23	46	49
30.5	62	53	23.5	50	50
34.5	65	54	35	54	51
35.5	73	56	26.5	58	52
37	77	57	27	62	53
37.5	81	59	28	65	54
39.5	85	60	29.5	73	56
41	92	65	31	77	57
43.5	96	67	35	81	59
50	99	72	36	89	62
			37	92	65
			39.5	96	67
			46.5	99	72

N = 26

**Appendix Table 9.**

Normative values (T-score and %ile) for raw scores for Carson-Jones Thematic Scoring [Age range 80–89; Education 17–20 years].

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
16.5	11	37	13	11	37
21	22	42	16.5	22	42
23.5	33	45	22.5	33	45
31.5	56	51	24	44	48
35.5	67	54	24.5	56	51
36	89	62	26	67	54

Carson-Jones Thematic Immediate			Carson Jones Thematic Delay		
raw	%ile	T-score	raw	%ile	T-score
47.5	99	72	26.5	78	57
			33	89	62
			45	99	72
N = 9					

## References

- Beck AT, Steer RA, & Brown GK (1996). Beck depression inventory-II. *San Antonio*, 78(2), 490–8.
- Benton AL, Hamsher KD, Varney NR, & Spreen O (1983). Judgment of line orientation. New York: Oxford University Press.
- Brandt J, & Benedict RHB (1991). Hopkins Verbal Learning Test-Revised (HVLTR). Lutz, FL: Psychological Assessment Resources.
- Carlesimo GA, Piras F, Orfei MD, 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: Diagnosis, Assessment & Disease Monitoring*, 1(1), 24–32.
- Chelune GJ, Naugle RI, Lüders H, Sedlak J, & Awad IA (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, 7(1), 41–52. doi: 10.1037/0894-4105.7.1.41
- Chételat G, Fouquet M, Kalpouzos G, Degenhien I, De La Sayette V, Viader F,...& Desgranges B (2008). Three-dimensional surface mapping of hippocampal atrophy progression from MCI to AD and over normal aging as assessed using voxel-based morphometry. *Neuropsychologia*, 46(6), 1721–1731. [PubMed: 18289618]
- Christianson SÅ, & Loftus EF (1991). Remembering emotional events: The fate of detailed information. *Cognition & Emotion*, 5(2), 81–108.
- Craft S, Newcomer J, Kanne S, Dagogo-Jack S, Cryer P, Sheline Y,...& Alderson A (1996). Memory improvement following induced hyperinsulinemia in Alzheimer's disease. *Neurobiology of aging*, 17(1), 123–130. [PubMed: 8786794]
- Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D,...Killiany RJ, 2006 An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. [PubMed: 16530430]
- Eysenck MW (1976). Arousal, learning, and memory. *Psychological Bulletin*, 83(3), 389. [PubMed: 778883]
- Fischl B (2012). FreeSurfer. *Neuroimage*, 62, 774–781. [PubMed: 22248573]
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C,...& Dale AM (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33, 341–355.10.1016/S0896-6273(02)00569-X [PubMed: 11832223]
- Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DHB,..., Dale AM, 2004 Automatically parcellating the human cerebral cortex. *Cerebral Cortex* 14, 11–22. [PubMed: 14654453]
- Folstein MF, Folstein SE, & McHugh PR (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12(3), 189–198. [PubMed: 1202204]
- Gavett BE, Ashendorf L, & Gurnani AS (2015). Reliable change on neuropsychological tests in the Uniform Data Set. *Journal of the International Neuropsychological Society*, 21(7), 558–567. [PubMed: 26234918]

- Gavett BE, Gurnani AS, Saurman JL, Chapman KR, Steinberg EG, Martin B,...Stern RA (2016). Practice effects on story memory and list learning tests in the neuropsychological assessment of older adults. *PLoS ONE*, 11(10).
- Golden CJ, & Freshwater SM (2002). *Stroop Color and Word Test, Revised 2002 Adult Manual for Clinical and Experimental Uses*. Wood Dale, IL: Stoelting.
- Heuer F, & Reisberg D (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & cognition*, 18(5), 496–506. [PubMed: 2233262]
- Iglesias JE, Augustinack JC, Nguyen K, Player CM, Player A, Wright M,... Van Leemput K (2015). A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI. *Neuroimage*, 115, 117–137. [PubMed: 25936807]
- Jacobson NS, & Truax P (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of consulting and clinical psychology*, 59(1), 12. [PubMed: 2002127]
- Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, & Delis DC (2009). Quantification of five neuropsychological approaches to defining mild cognitive impairment. *The American Journal of Geriatric Psychiatry*, 17(5), 368–375. [PubMed: 19390294]
- Klein A, Tourville J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience*, 6, 171. [PubMed: 23227001]
- La Joie R, Fouquet M, Mézence F, Landeau B, Villain N, Mevel K,...Chételat G (2010). Differential effect of age on hippocampal subfields assessed using a new high-resolution 3T MR sequence. *Neuroimage*, 53(2), 506–514. [PubMed: 20600996]
- Larrabee GJ, Kane RL, & Schuck JR (1983). Factor analysis of the WAIS and Wechsler Memory Scale: An analysis of the construct validity of the Wechsler Memory Scale. *Journal of Clinical and Experimental Neuropsychology*, 5(2), 159–168.
- Lawton MP, & Brody EM (1969). Assessment of older people: self-maintaining and instrumental activities of daily living. *The gerontologist*, 9(3\_Part\_1), 179–186. [PubMed: 5349366]
- Leonberger FT, Nicks SD, Larrabee GJ, & Goldfader PR (1992). Factor structure of the Wechsler Memory Scale—Revised within a comprehensive neuropsychological battery. *Neuropsychology*, 6(3), 239.
- Lezak M, Howieson DB, Loring DW. (2004). *Neuropsychological Assessment* (4 th ed.). New York: Oxford University Press.
- McCaffrey RJ, Ortega A, Orsillo SM, Nelles WB, & Haase RF. (1992). Practice effects in repeated neuropsychological assessments. *The Clinical Neuropsychologist*, 6(1), 32–42.
- McCaffrey RJ, Ortega A, & Haase RF. (1993). Effects of repeated neuropsychological assessments. *Archives of Clinical Neuropsychology*. 10, 241–250.
- McCaffrey R, Duff K, & Westervelt H (2000). *Practitioner's Guide to Evaluating Change with Neuropsychological Assessment Instruments*. New York: Kluwer Academic/Plenum Press.
- Morris J, Kunka JM, & Rossini ED. (1997). Development of alternate paragraphs for Logical Memory subtest of the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist*. 11, 370–374.
- Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I,...& 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. [PubMed: 15817019]
- Newcomer JW, Craft S, Hershey T, Askins K, & Bardgett ME. (1994). Glucocorticoid-induced impairment in declarative memory performance in adult humans. *The Journal of Neuroscience*. 14(4), 2047–2053. [PubMed: 8198631]
- Newcomer JW, Craft S, Askins K, Hershey T, Bardgett ME, Csernasky JG,...& Vogler G (1997). Glucocorticoid interactions with memory function in schizophrenia. *Psychoendocrinology*. 23(1), 65–72.
- Newcomer JW, Craft S, Fucetola R, Moldin SO, Selke G, Paras L, & Miller R (1999). Glucose-induced increase in memory performance in patients with schizophrenia. *Schizophrenia Bulletin*. 25(2), 321–335. [PubMed: 10416734]

- Newcomer JW, Selke G, Melson AK, Hershey T, Craft S, Richards K, & Alderson AL. (1999). Decreased memory performance in healthy humans induced by stress-level cortisol treatment. *Arch Gen Psychiatry*. 56, 527–533. [PubMed: 10359467]
- Price CC, Wood MF, Leonard CM, Towler S, Ward J, Montijo H, ... & Schmalfluss I (2010). Entorhinal cortex volume in older adults: Reliability and validity considerations for three published measurement protocols. *Journal of the International Neuropsychological Society*, 16(05), 846–855. [PubMed: 20937164]
- Schnabel R (2012). Overcoming the challenge of re-assessing logical memory. *The Clinical Neuropsychologist*. 26(1), 102–115. [PubMed: 22172088]
- Salthouse T, Schroeder D, & Ferrer E (2004). Estimating retest effects in longitudinal assessments of cognitive functioning in adults between 18 and 60 years of age. *Developmental Psychology*. 40, 813–822. [PubMed: 15355168]
- Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, & Jacobs GA (1983). *State-trait anxiety inventory (form Y)*. Redwood City, CA: Mind Garden, 77.
- Squire LR. (1987). *Memory and brain*. New York: Oxford University Press.
- Squire LR & Zola-Morgan S (1991). The medial temporal lobe memory system. *Science*. 253, 1380–1386. [PubMed: 1896849]
- Squire LR. (1992). Memory and the hippocampus. A synthesis from findings with rats, monkeys, and humans. *Psychological Review*. 99, 195–231. 20(6), 745–753. [PubMed: 1594723]
- Starkstein SE, Mayberg HS, Preziosi T, Andrezejewski P, Leiguarda R, & Robinson RG (1992). Reliability, validity, and clinical correlates of apathy in Parkinson's disease. *J Neuropsychiatry Clin Neurosci*. 4(2), 134–139. [PubMed: 1627973]
- Sullivan K (2005). Alternate forms of prose passages for the assessment of auditory-verbal memory. *Archives of Clinical Neuropsychology*.
- Theisen ME, Rapport LJ, Axelrod BN, & Brines DB (1998). Effects of practice in repeated administrations of the Wechsler memory scale-revised in normal adults. *Assessment*, 5(1), 85–92. [PubMed: 9458344]
- Tulsky DS, & Price LR (2003). The joint WAIS-III and WMS-III factor structure: development and cross-validation of a six-factor model of cognitive functioning. *Psychological assessment*, 15(2), 149. [PubMed: 12847775]
- Wechsler D (1945). A standardized memory scale for clinical use. *The Journal of Psychology*. 19, 87–95.
- Wechsler D (2001). *Wechsler Test of Adult Reading: WTAR*. Psychological Corporation.
- Wilkinson GS & Robertson GJ (2006). *Wide Range Achievement Test 4: WRAT-4*. Psychological Assessment Resources.
- Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, & Leirer VO (1982). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of psychiatric research*, 17(1), 37–49. [PubMed: 7183759]

---

**Lucy / Carson, / while / visiting / her / brother / in a small / city / in Virginia, / got lost / walking / to the art / museum. / She / searched / her / pockets / for the directions / that / had been written / for her. / Unfortunately, / she / had left / them / sitting / on her / wooden / dresser. / She / then / asked / a bus / driver / for directions; / he / said / to / walk / north / on 11<sup>th</sup> / Avenue / for five / blocks. /**

Prompt: “A woman who went on a trip.”

---

**Figure 1.**  
Lucy Carson Story (Carson-Jones Story Set)

---

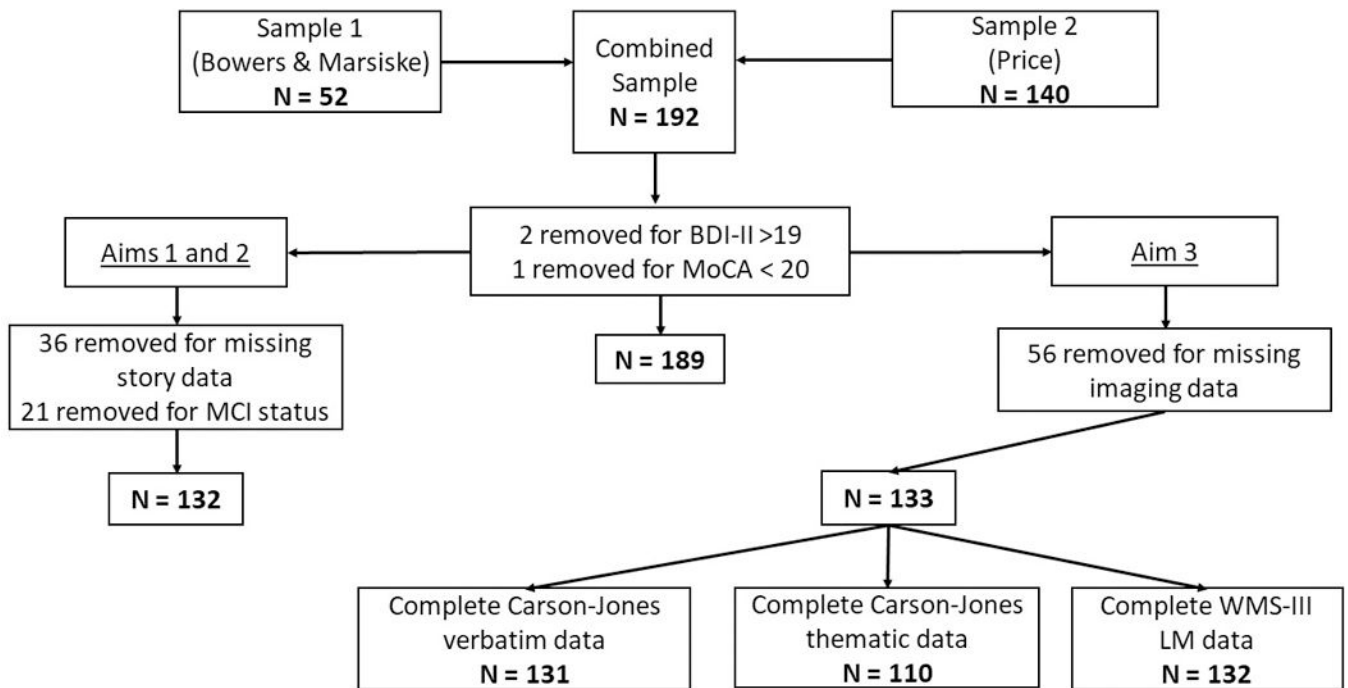
**Adam / Jones' / department / store / held / an annual / sale / Tuesday / through  
the end / of the week. / Sporting / goods / were marked / down / 75 / percent, /  
and winter / jackets / were reduced / by 30 / dollars. / Crowds / of shoppers /  
bumped / each / other, / grabbing / for the best / buys. / After / closing / time, /  
when the sale / was / over, / the owner / gave / each / employee / a bottle / of  
champagne / for their / hard / work. /**

Prompt: "A place that sold things."

---

**Figure 2.**  
Adam Jones Story (Carson-Jones Story Set)





**Figure 3.** Sample Size Composition Flow Chart for Each Aim. BDI-II = Beck Depression Inventory, 2<sup>nd</sup> Edition; MoCA = Montreal Cognitive Assessment; WMS-III LM = Wechsler Memory Scale, 3<sup>rd</sup> Edition, Logical Memory subtest; MCI = mild cognitive impairment

**Table 1.**

Demographics and mood characteristics of psychometrically normal participants from Sample 1 and Sample 2

	Sample 1 ( <i>n</i> = 44)			Sample 2 ( <i>n</i> = 88)		
	Mean ( <i>SD</i> )	Min	Max	Mean ( <i>SD</i> )	Min	Max
<b>Age (yrs)</b>	76 (7.5) *	61	92	69 (6.2) *	60	85
<b>Race (%White)</b>	98	-	-	97	-	-
<b>Education (yrs)</b>	17 (2.1) *	12	20	16 (2.5) *	9	20
<b>Gender (%Male)</b>	29	-	-	44	-	-
<b>Handedness (%Right)</b>	93	-	-	91	-	-
<b>MoCA *</b>	-	-	-	25 (2.5)	14	29
<b>MMSE *</b>	29 (1.0)	27	30	-	-	-
<b>CCI</b>	0.61 (1.5)	0	8	0.31 (0.58)	0	2
<b>BDI-II</b>	5 (3.9)	0	18	5 (4.8)	0	17
<b>GDS</b>	3 (2.8)	0	8	3 (3.6)	0	16
<b>AS</b>	9 (4.8)	0	23	8 (3.7)	0	18
<b>STAI-Trait</b>	28 (7.4)	20	51	29 (7.4)	20	50

\* = significantly different at  $p < .05$ ;MoCA = Montreal Cognitive Assessment; MMSE = Mini-Mental Status Exam; CCI = Charlson Comorbidity Index; BDI-II = Beck Depression Inventory, 2<sup>nd</sup> Edition; GDS = Geriatric Depression Scale; AS = Apathy Scale; STAI-Trait = State-Trait Anxiety Inventory, Trait

**Table 2.**Demographics and mood characteristics of combined sample:  $n=132$  psychometrically normal participants

	Mean (SD)	Min	Max
<b>Age (yrs)</b>	72 (7.5)	60	92
<b>Education (yrs)</b>	16 (2.4)	9	20
<b>Race (%White)</b>	97	-	-
<b>Gender (%Male)</b>	39	-	-
<b>Handedness (%Right)</b>	91	-	-
<b>CCI</b>	0.41 (1.0)	0	8
<b>BDI-II</b>	4.8 (4.5)	0	18
<b>GDS</b>	3 (3.3)	0	16
<b>AS</b>	8.6 (4.1)	0	23
<b>STAI-Trait</b>	29 (7.4)	20	51

CCI = Charlson Comorbidity Index; BDI-II = Beck Depression Inventory, 2<sup>nd</sup> Edition; GDS = Geriatric Depression Scale; AS = Apathy Scale; STAI-Trait = State-Trait Anxiety Inventory, Trait

**Table 3.**

Correlation matrix between Carson-Jones (Newcomer) story memory and other measures of interest:  $n=132$  psychometrically normal participants

	LM I	LM II	Carson-Jones Imm Verbatim	Carson-Jones Delay Verbatim	Carson-Jones Imm Thematic	Carson-Jones Delay Thematic
<b>LM I</b>	-	-	-	-	-	-
<b>LM II</b>	0.809 **	-	-	-	-	-
<b>Carson-Jones Imm - V</b>	0.665 **	0.587 **	-	-	-	-
<b>Carson-Jones Delay - V</b>	0.651 **	0.633 **	0.847 **	-	-	-
<b>Carson-Jones Imm - T</b>	0.642 **	0.582 **	0.909 **	0.846 **	-	-
<b>Carson-Jones Delay - T</b>	0.671 **	0.630 **	0.851 **	0.895 **	0.920 **	-
<b>HVLT Trials 1-3</b>	0.525 **	0.595 **	0.472 **	0.474 **	0.484 **	0.464 **
<b>HVLT Delay</b>	0.538 **	0.576 **	0.436 **	0.402 **	0.432 **	0.422 **
<b>Stroop Word</b>	0.210 *	0.229 **	0.044	0.107	0.019	0.068
<b>JOLO</b>	-0.089	-0.008	-0.163	-0.133	-0.114	-0.113

\*\* = significant at  $p < .001$ ;

\* = significant at  $p < .05$ ;

each story variable represents total combined story scores (i.e., LM = Story A + B; Carson-Jones = Story 1 + 2); LM I = Logical Memory Immediate; LM II = Logical Memory Delay; HVLT = Hopkins Verbal Learning Test; JOLO = Judgment of Line Orientation

**Table 4.**

Multiple regression equations with story memory scores predicting left hemisphere structural variables (with both psychometrically healthy and mild cognitive impairment included)

		Left Subiculum Volume	Left Whole Hippocampus Volume	Left Entorhinal Cortex Thickness	Left Rostral Middle Frontal Thickness
<b>Carson-Jones Delay - V</b>	F	10.302 <sup>**</sup>	5.399 <sup>*</sup>	0.255	1.040
	R <sup>2</sup>	0.062	0.032	0.002	0.007
<b>Carson-Jones Delay - T</b>	F	4.284 <sup>*</sup>	2.389	0.294	1.739
	R <sup>2</sup>	0.033	0.018	0.003	0.015
<b>LM Delay</b>	F	6.241 <sup>*</sup>	4.187 <sup>*</sup>	3.604 <sup>t</sup>	0.088
	R <sup>2</sup>	0.038	0.021	0.026	0.001

<sup>\*\*</sup> = significant at  $p < .01$ ;

<sup>\*</sup> = significant at  $p < .05$ ;

<sup>t</sup> = trend significance at  $p < .10$ ;

Carson-Jones Delay – Verbatim  $n = 131$ ; Carson-Jones Delay – Thematic  $n = 110$ ; LM II  $n = 132$ ; LM = Logical Memory; Carson-Jones Delay – V = Carson-Jones Delayed Recall Verbatim; Carson-Jones Delay – T = Carson-Jones Delay Thematic

**Table 5.**

Multiple regression equations with story memory scores predicting right hemisphere structural variables (with both psychometrically healthy and mild cognitive impairment included)

		<b>Right Subiculum Volume</b>	<b>Right Whole Hippocampus Volume</b>	<b>Right Entorhinal Cortex Thickness</b>	<b>Right Rostral Middle Frontal Thickness</b>
<b>Carson-Jones Delay - V</b>	F	2.347	2.499	0.217	0.013
	R <sup>2</sup>	0.016	0.016	0.002	0.000
<b>Carson-Jones Delay - T</b>	F	1.787	1.448	0.105	0.010
	R <sup>2</sup>	0.015	0.012	0.001	0.000
<b>LM Delay</b>	F	2.864 <sup>t</sup>	2.366	0.153	0.352
	R <sup>2</sup>	0.019	0.015	0.001	0.003

<sup>t</sup> = trend significance at  $p < .10$ ;

Carson-Jones Delay – Verbatim  $n = 131$ ; Carson-Jones Delay – Thematic  $n = 110$ ; LM II  $n = 132$ ; LM = Logical Memory; Carson-Jones Delay – V = Carson-Jones Delay Verbatim; Carson-Jones Delay – T = Carson-Jones Delay Thematic