Routledge
Taylor & Francis Group

Aging, Neuropsychology, and Cognition

A Journal on Normal and Dysfunctional Development

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/nanc20

Sensitivity of memory subtests and learning slopes from the ADAS-Cog to distinguish along the continuum of the NIA-AA Research Framework for Alzheimer's Disease

Dustin B Hammers, Ralitsa V Kostadinova, Robert J Spencer, Jean N Ikanga, Frederick W Unverzagt, Shannon L Risacher, Liana G Apostolova & for the Alzheimer's Disease Neuroimaging Initiative

To cite this article: Dustin B Hammers, Ralitsa V Kostadinova, Robert J Spencer, Jean N Ikanga, Frederick W Unverzagt, Shannon L Risacher, Liana G Apostolova & for the Alzheimer's Disease Neuroimaging Initiative (2022): Sensitivity of memory subtests and learning slopes from the ADAS-Cog to distinguish along the continuum of the NIA-AA Research Framework for Alzheimer's Disease, Aging, Neuropsychology, and Cognition, DOI: 10.1080/13825585.2022.2120957

To link to this article: https://doi.org/10.1080/13825585.2022.2120957

Published online: 08 Sep 2022.

AGING, NEUROPSYCHOLOGY, AND COGNITION
https://doi.org/10.1080/13825585.2022.2120957

Sensitivity of memory subtests and learning slopes from the ADAS-Cog to distinguish along the continuum of the NIA-AA Research Framework for Alzheimer's Disease

Dustin B Hammers[a], Ralitsa V Kostadinova[a], Robert J Spencer[b,c], Jean N Ikanga[d,e], Frederick W Unverzagt[f], Shannon L Risacher[g], Liana G Apostolova[a] and for the Alzheimer's Disease Neuroimaging Initiative

[a]Department of Neurology, Indiana University School of Medicine, Indianapolis, IN, USA; [b]Mental Health Service, VA Ann Arbor Healthcare System, Ann Arbor, MI, USA; [c]Michigan Medicine, Department of Psychiatry, Neuropsychology Section, Ann Arbor, MI, USA; [d]Department of Rehabilitation Medicine, Emory University, School of Medicine, Atlanta, GA, USA; [e]Department of Psychiatry, Democratic Republic of Congo (DRC), University of Kinshasa; [f]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA; [g]Department of Radiology, Indiana University School of Medicine, Indianapolis, IN, USA

ABSTRACT

Despite extensive use of the Alzheimer's Disease (AD) Assessment Scale – Cognitive Subscale (ADAS-Cog) in AD research, exploration of memory subtests or process scores from the measure has been limited. The current study sought to establish validity for the ADAS-Cog Word Recall Immediate and Delayed Memory subtests and learning slope scores by showing that they are sensitive to AD biomarker status. Word Recall subtest and learning slope scores were calculated for 441 participants from the Alzheimer's Disease Neuroimaging Initiative (aged 55 to 90). All participants were categorized using the NIA-AA Research Framework – based on PET-imaging of β-amyloid (A) and tau (T) deposition – as Normal AD Biomarkers (A-T-), Alzheimer's Pathologic Change (A + T-), or Alzheimer's disease (A + T+). Memory subtest and learning slope performances were compared between biomarker status groups, and with regard to how well they discriminated samples with (A + T +) and without (A-T-) biomarkers. Lower Word Recall memory subtest scores – and scores for a particular learning slope calculation, the Learning Ratio – were observed for the AD (A + T+) group than the other biomarker groups. Memory subtest and Learning Ratio scores further displayed fair to good receiver operator characteristics when differentiating those with and without AD biomarkers. When comparing across learning slopes, the Learning Ratio metric consistently outperformed others. ADAS-Cog memory subtests and the Learning Ratio score are sensitive to AD biomarker status along the continuum of the NIA-AA Research Framework, and the results offer criterion validity for use of these subtests and process scores as unique markers of memory capacity.

ARTICLE HISTORY
Received 6 May 2022
Accepted 30 August 2022

KEYWORDS
Learning; memory; alzheimer's disease; mild cognitive impairment; tau; amyloid; ADAS-Cog

CONTACT Dustin B Hammers HammersD@iu.edu

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Introduction

The Alzheimer's Disease (AD) Assessment Scale – Cognitive Subscale (ADAS-Cog) has been used extensively since its creation nearly 40 years ago (Rosen et al., 1984). Its frequent use in AD pharmaceutical intervention trials (Birks, 2006) has led to it being considered the gold standard of cognitive assessment in AD clinical trials (Kueper et al., 2018). The updated version of the measure (Mohs et al., 1997) is comprised of 13 subtests that assess a wide range of cognitive abilities (including delayed memory), with a total score of 85 points where higher scores reflect worse performance. The association between performance on the ADAS-Cog and AD biomarkers has been strong, with a recent meta-analysis of 17 studies suggesting that average performance on the ADAS-Cog correlated with amyloid-Positron Emission Tomography (PET) standardized uptake value ratios (SUVRs) at r = .68 (R^2 = .46; Avgerinos et al., 2021). Similarly, ADAS-Cog performance correlates strongly with tau-PET SUVRs (r = .59, R^2 = .51; Devous et al., 2021). Further, a meta-analysis of 60 studies indicated that – along with amyloid and tau cerebrospinal fluid levels – hippocampal/medial temporal lobe atrophy and ADAS-Cog performance were two of the strongest predictors of progression from Mild Cognitive Impairment (MCI) to AD (Li et al., 2016).

Despite the widespread use of the ADAS-Cog in AD research, investigation of individual subtests that make up the Total Score has been relatively rare. This is likely based on the prevailing sense that ADAS-Cog subtests are generally insensitive, particularly at identifying impairments in less-severe populations (Wesnes, 2008). Using item-response theory analysis, Benge et al. (2009) found that for most ADAS-Cog subscales, sensitivity was greatest at moderate levels of cognitive dysfunction – which appears to support this prevailing opinion. However, they additionally observed that performance on memory subtests discriminated best at lower levels of cognitive dysfunction, which is a finding that has been later supported by work in participants with MCI (Lowe et al., 2015; Sano et al., 2011) and subjective memory impairment (Lowe et al., 2015). The ADAS-Cog contains a Word Recall (Immediate Recall) subtest that is comprised of a 10-item word-list that is repeated successively over three trials, along with a Delayed Recall subtest after a ~10 minute delay (Mohs et al., 1997). As memory subtests from the ADAS-Cog appear to be sensitive to mild-to-moderate cognitive impairments, it is likely that – like the ADAS-Cog Total Score – these subtests may also be sensitive to changes along the AD biomarker continuum. However, we have been unable to identify any literature that examines performance of memory subtests of the ADAS-Cog in the context of the recently created biomarker-based National Institutes of Health-Alzheimer's Association (NIA-AA) Research Framework for AD (Jack et al., 2018).

Similarly, investigation of learning slope process scores in the ADAS-Cog has yet to be undertaken. As markers of learning slope have been shown to be deficient in patients with AD (Gifford et al., 2015; Hammers, Suhrie et al., 2021) and related to amyloid burden (Hammers, Suhrie et al., 2021), learning slopes derived from the memory subtests from the ADAS-Cog may relatedly show sensitivity along the biomarker spectrum. Multiple methods exist to calculate learning slopes, which will be described in more detail in the Methods. Briefly, the most common metric is the Raw Learning Score (RLS), which reflects knowledge acquired over trials of a multi-trial learning task – and is calculated as the difference between the first learning trial and

the final learning trial (Bender et al., 2020; Brandt & Benedict, 1997). Also, Learning Over Trials (LOT) subtracts the first trial value from each subsequent trial of a multi-trial learning task, and reflects incremental learning after removing the impact of Trial-1 performance (Morrison et al., 2018; Thomas et al., 2020). Further, the Learning Ratio (LR) considers that most learning slope calculations are confounded by the amount of available information yet-to-be-learned after the first trial of a multi-trial learning task, and therefore factors unlearned information into its calculus. For example, if Participant 1 learns 3 words on a 3-trial 10-item word-list at Trial 1 and 6 words at Trial 3, and Participant 2 learns 7 words at Trial 1 and 10 words at Trial 3, they would have each learned 3 words over trials (6–3 = 3; 10–7 = 3); while this RLS score would be comparable, it fails to acknowledge Participant 2's stronger overall learning acquisition, and that Participant 2 only acquired 3 more items because *that was all that was left to learn*. Instead, LR takes the basic RLS calculation (final trial minus first trial) and divides by number of items that are still available to learn after the first trial (Spencer et al., 2020). In our example above, Participant 1 had 7 words left-to-learn after Trial 1, and Participant 2 had 3. Consequently, LR – or the proportion of yet-to-be-learned information – for Participant 1 was 43% (3/7 = 0.43 * 100), and for Participant 2 was 100% (3/3 = 1.0 * 100). This difference between learning slope outcomes appears to fit more closely to our perceptions of learning capacities for these two participants and raises the possibility that LR may be particularly sensitive to differences in AD biomarker status.

The primary aim of the current study is to examine the sensitivity of the memory subtests from the ADAS-Cog – and the learning slope metrics derived from them – at detecting differences in AD biomarker status using the recently developed NIA-AA Research Framework for the diagnosis of AD (Jack et al., 2018). The limited research into both ADAS-Cog memory subtests and learning slopes among individuals along the AD biomarker continuum represents a gap in the literature. Given the prevalence of use of the ADAS-Cog in AD research, and the elevated cost of either biomarker imaging or CSF assays (Wolk et al., 2019), identifying subtests and/or process scores (learning slopes) from commonly administered cognitive measures that are sensitive to AD biomarker status represents an opportunity to enhance the fields' capacity to cost-effectively predict the presence of AD pathology. Consequently, data from participants enrolled in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (Weiner et al., 2017) who underwent cognitive assessment and both β-amyloid- PET and tau-PET scans were used in the present study. Initially, memory and learning slope performance from participants across the AD biomarker continuum were compared, with the hypothesis being that participants with more severe AD biomarker profiles will perform worse cognitively. Additionally, we examined the capacity for the various memory subtests and learning slope metrics to discriminate between individuals without AD biomarkers and those with AD, and we hypothesized that memory and learning slope metrics would possess adequate to good test receiver operator characteristics (e.g., area under the curve, specificity, sensitivity). Further, given evidence that LR may be superior to other learning slopes at predicting amyloid-positivity (Hammers, Suhrie, Dixon, Gradwohl, Archibald et al., 2021), in each of the above analyses we supplementarily compared the learning slopes – LR, RLS, and LOT – to each other, with the hypothesis being that LR would better discriminate AD biomarker status than the other learning slopes. Overall, should our hypotheses be correct, our results would provide support for the sensitivity of both memory subtests and learning

slope metrics from the ADAS-Cog toward AD biomarker status, and potentially suggest a cost-effective tool to aid in the detection of AD pathology.

Methods

The participant data for the current study were obtained from ADNI's multi-center longitudinal study (http://adni.loni.usc.edu). As has been described elsewhere (Weiner et al., 2017), ADNI is as a public–private partnership launched in 2003 with the primary goal of testing whether serial magnetic resonance imaging (MRI), PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org. Institutional Review Board approval for ADNI has been obtained for each multi-center sites, and written informed consent was obtained from study participants or authorized representatives.

Cognitive data were available for 2366 ADNI participants across various ADNI protocols (ADNI1, 2004; ADNI2., 2011; ADNI3., 2016; ADNIGO, 2009) as of April 26th, 2021. Participant data collection began on 08/23/2005 and enrolled participants have been followed cognitively for up to 180 months. Tau-PET investigation was initiated in 2017. Inclusion for ADNI involved the following: having ≥6 years of education and having a reliable study partner; being between the ages of 55 to 90 at baseline; absence of significant head trauma, depression, or neurologic disease; stability on permitted medications; and fluency in either English or Spanish (ADNI1, 2004; ADNI2., 2011; ADNI3., 2016; ADNIGO, 2009). For the current study, 1,868 participants were excluded for not having Tau-PET data from their baseline visit. An additional 43 participants were excluded for having missing baseline cognitive or amyloid-PET data, resulting in 456 participants who had relevant cognitive data and were able to be classified using the NIA-AA Research Framework based on PET-imaging of β-amyloid (A) and tau (T) deposition at their baseline visit. Please see, Figure 1 for a schematic representation of the current study's participant utilization from ADNI.

The classification of β-amyloid and tau positivity from the baseline visit was conducted as follows:

β-amyloid Positivity

Participants in the current study underwent amyloid-PET imaging using the radioligand ^{11}C-labeled Pittsburgh Compound B, ^{18}F-Florbetapir, or ^{18}F-Florbetaben, as per ADNI protocols (ADNI1, 2004; ADNI2., 2011; ADNI3., 2016; ADNIGO, 2009). Radioligand deposition was calculated as a whole-brain SUVR, normalized to whole cerebellum. SUVR cutoffs for the respective radioligands were ≥1.50, ≥1.11, and ≥1.08 for ^{11}C-labeled Pittsburgh Compound B, ^{18}F-Florbetapir, or ^{18}F-Florbetaben, respectively. Please see ADNI protocols for greater details about amyloid-PET methods in ADNI.

Tau-Positivity

All participants in the current study underwent tau-PET imaging using ^{18}F-Flortaucipir, as per ADNI protocols (ADNI3., 2016). Tau-positivity was classified using Schwarz and colleagues' (Schwarz et al., 2018) Tau-PET topographical staging scheme for AD. Specifically,

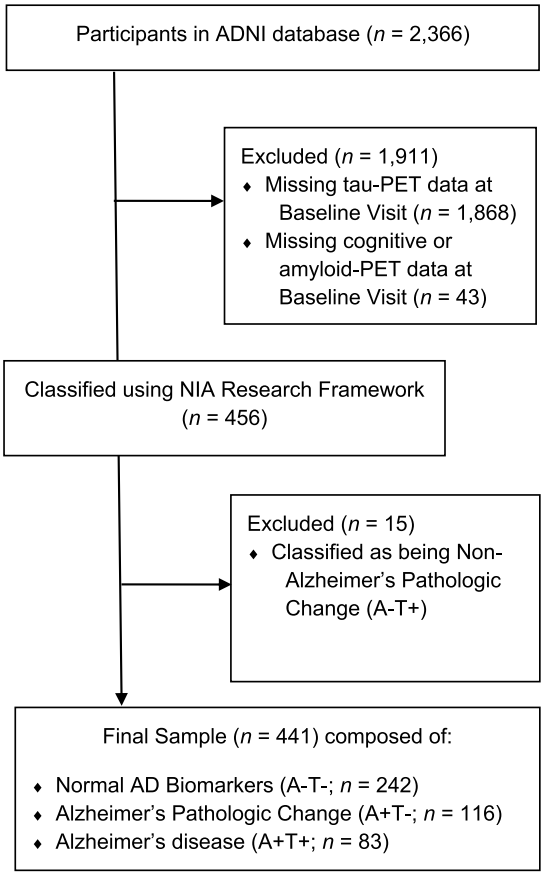

Figure 1. Flow diagram of participants recruited into the current study from the total sample of ADNI participants.

pre-defined patterns of tau-burden were identified from the following brain regions based on the associated SUVR cutoff (in parentheses): hippocampus (SUVR threshold of ≥1.222), transentorhinal cortex (≥1.310), fusiform gyrus (≥1.352), middle temporal gyrus (≥1.296), superior temporal gyrus (≥1.219), extrastriate visual cortex (≥1.308), and primary visual cortex (≥1.268). Resultant patterns of positivity led to a Pathologic Staging score ranging from 0 to 6, with scores of 0–3 being considered tau negative, and scores of 4–6 being considered tau positive.

Following the application of amyloid and tau classifications, we designated participants into established AD biomarker categories using the NIA-AA Research Framework (Jack et al., 2018). Of note, while the full Research Framework uses a combination of amyloid-positivity (A), tau-positivity (T), and neurodegeneration-positivity (N) to classify participants (i.e., "AT(N)"), the current study only incorporated data from "A and T." This is because biomarkers of neurodegeneration have limited impact on the AD biomarker categories and are considered ancillary given their known limitations at diagnosing AD, including (1) the non-specificity of neurodegeneration toward a particular etiology (like AD), and (2) the failure of neurodegenerative biomarkers to map onto

neuropathologic findings used to diagnose AD (Jack et al., 2018). The results of classification led to 242 participants being categorized as having Normal AD Biomarkers (A-T-), 116 participants being categorized as Alzheimer's Pathologic Change (A + T-), 15 participants being categorized as Non-Alzheimer's Pathologic Change (A-T+), and 83 participants being categorized as Alzheimer's disease (A + T+). Given the infrequency of the A-T + participants, and their status as being likely outside the Alzheimer's continuum (Jack et al., 2018), they were excluded from the analyses. Consequently, as seen in Figure 1, the final sample reflected 441 participants with Normal AD Biomarkers (A-T-), Alzheimer's Pathologic Change (A + T-), or Alzheimer's disease (A + T+).

For further characterization of the AD biomarker status groups, participants were classified into diagnostic groups using ADNI criteria. The reader is referred to previous documentation of ADNI's classification criteria (ADNI3., 2016), though briefly Logical Memory from the Wechsler Memory Scale – Revised (WMS-R; Wechsler, 1987), the Mini-Mental State Examination (MMSE; Folstein et al., 1975), and the Clinical Dementia Rating (CDR) scale (Morris, 1993) were used to determine diagnostic classifications. As such, neither the ADAS-Cog Word Recall, Rey Auditory Verbal Learning Test (RAVLT; Schmidt, 1996), nor learning slopes were used to inform diagnosis in this sample.

Procedure

All participants underwent an extensive clinical and neuropsychological battery at a baseline visit upon their enrollment in ADNI. Although the 15-item Geriatric Depression Scale (Sheikh & Yesavage, 1986; cutoff for depression >5), CDR, and Montreal Cognitive Assessment (MOCA; Nasreddine et al., 2005) are used in the current study as demographic descriptors, the reader is referred to ADNI protocols (ADNI3., 2016) or test developer manuals for further description of these measures. For the current study, the relevant neuropsychological measures used were as follows:

- The ADAS-Cog (Rosen et al., 1984) is a neuropsychological test battery comprising 13 subtests pertaining to learning and memory, language production and comprehension, constructional praxis, ideational praxis, orientation, and executive skills. The range of scores for the ADAS-Cog Total Score is from 0 to 85, with higher scores indicating *worse* performance. Of the ADAS-Cog subtests, the Word Recall subtest (Question 1 of the ADAS-Cog) is a verbal memory task with 10 words learned over 3 trials, and words from this list cannot easily be clustered into semantic categories. The Delayed Recall subtest (Question 4 of the ADAS-Cog) requests participants to recall those words after a 10-min delay. For the purpose of the current study, the Immediate Recall score is the number of correct words identified across trials (range = 0–30), and the Delayed Recall score is the number of correct words recalled after delay (range = 0–10). This subtest scoring deviates from test developer's protocols, as the original scoring aims to show worse performance resulting in *higher* scores (i.e., the number of words not learned/recalled). However, this alternative scoring was instituted for consistency with all other memory measures in the study; as a result, higher values for ADAS-Cog Word Recall Immediate Recall and Delayed Recall in our study indicate *better* performance. Learning slope performances were evaluated by raw data from individual Word Recall trials and will be described later in the Methods.

●RAVLT is a verbal memory task with 15 words learned over 5 trials, with the number of correct words summed for the Total Recall score (range = 0–75). The Delayed Recall score is the number of correct words recalled after a 30-min delay (range = 0–15). All RAVLT scores are raw scores, with higher values indicating better performance.

●Logical Memory I and II from the WMS-R (Wechsler, 1987) are immediate and delayed (20–30 minute) memory measures assessing capacity to acquire and retain information from a verbally presented short story. Of note, according to ADNI protocol, only "Story A" was provided to participants, therefore the range of scores for Logical Memory I and II is both 0–23. All values are raw scores, with higher values indicating better performance.

●American National Adult Reading Test (AMNART; Grober & Sliwinski, 1991) is an estimate of premorbid verbal intellect. For this measure, participants attempt to pronounce 45 words for which the pronunciation does not follow common phonetic rules. The total number of errors made is entered into a regression equation with years of education to yield the estimate of verbal intelligence in standard scores (M = 100, SD = 15). Higher values indicate higher estimated baseline intellectual functioning.

Calculation of Learning Slopes

For the Word Recall subtest of the ADAS-Cog, RLS scores were computed as performance on Trial 3 minus Trial 1. LOT scores were computed as the sum of Trials 1 through 3 minus the value of Trial 1 multiplied by 3 (modified from Morrison et al., 2018). The LR score is represented as a proportion as follows: the difference in performance between Trial 3 and Trial 1 in the numerator, and the difference between the maximum possible trial score and Trial 1 performance in the denominator (Spencer et al., 2020). Please note that the "Total Points Available for a Trial" for the Word Recall subtest is 10, though the equations below are written broadly to apply to LR values derived from both the Word Recall subtest and other learning measures. The formulas for RLS, LOT, and LR derived from the Word Recall subtest of the ADAS-Cog are as follows:

$$\text{RLS} = (\text{Trial 3 performance} - \text{Trial 1 performance})$$

$$\text{LOT} = (\text{Sum of performance on Trials 1 through 3} - (3 * \text{Trial 1 performance}))$$

$$\text{LR} = \frac{(\text{Trial 3 performance Trial 1 performance})}{(\text{Total Points Available for a Trial Trial 1 performance})}$$

Data Analysis

For demographic comparisons and to determine the appropriateness of covariates, one-way analyses of variance (ANOVA) were calculated between AD biomarker status groups (A-T-, A + T-, and A + T+) for continuous demographic variables (e.g., age, education, and verbal intellect), and chi-square analyses were calculated for categorical demographic variables (e.g., sex and ethnicity) in relation to biomarker status groups. For additional consideration of covariation, bivariate correlations were conducted between demographic variables (i.e., age, education, verbal intellect, sex, and ethnicity) and Word Recall subtests and LR values.

For the criterion validity primary analyses, multivariate analysis of covariance (MANCOVA) was conducted comparing AD biomarker status (A-T-, A + T-, A + T+) on

ADAS-Cog Word Recall Immediate and Delayed Recall subtests, learning slope performances (LR, RLS, and LOT), and ADAS-Cog Total Score. Subsequent one-way analyses of covariance (ANCOVA) analyses were conducted for individual cognitive measures within the omnibus test. For significant ANCOVA analyses, Bonferroni post-hoc comparisons were calculated among biomarker status group performances.

For consideration of test operating characteristics for ADAS-Cog Word Recall Immediate and Delayed Recall subtests, learning slope, and ADAS-Cog Total Scores, receiver operating characteristic area under the curve (ROC-AUC) analyses were conducted between participants in the A + T+ and the A-T- groups. For the interpretation of ROC-AUC values, the current study followed guidelines suggested by Hosmer and colleagues (Hosmer et al., 2013) of ROC-AUC values <0.600 being a "failure," values between 0.600 and 0.699 being "poor," values between 0.700 and 0.799 being "fair," values between 0.800 and 0.899 being "good," and values 0.900 or greater being "excellent." Cut scores for Word Recall subtests, learning slopes, and the ADAS-Cog were determined based on optimal sensitivity and specificity for detecting the presence of both β-amyloid and tau pathology.

Finally, given results suggesting differences in both criterion validity and test operating characteristics among learning slopes derived from the Word Recall memory subtest, supplementary convergent validity analyses were conducted. Specifically, partial correlations between individual learning slope performances and traditional memory measures (from ADAS-Cog Word Recall, RAVLT, and Logical Memory) were conducted across AD biomarker status groups and the total sample, and Fishers *r* to *z* transformations examined differences in correlations between slope metrics.

Measures of effect size were expressed as partial eta squared (η^2; MANCOVA/ANCOVA) values, Phi coefficients (X^2), and R^2 values (bivariate and partial correlations). Additionally, comparisons between AUC values were examined using 95% confidence intervals (CIs). To protect against multiple comparisons, a two-tailed alpha level was set at .01 for all primary analyses.

Results

Demographics and Memory Testing

The primary sample was composed of 441 participants undergoing amyloid-PET and tau-PET from ADNI. As seen in Table 1, the mean age of the total sample was 70.95 (*SD* = 7.1; range 55–90) years old and the sample averaged 16.46 (*SD* = 2.3; range 10–20) years of education. The sample of participants was slightly more female predominant (56.2% female) and the majority of participants were Caucasian (85.3%). Mean intellect at baseline was estimated to be high average according to the AMNART Verbal Intellect standard score (*M* = 118.52, *SD* = 9.9, range 85–131), and overall self-reported depression was low (*M* = 1.11, *SD* = 1.4) according to the 15-item Geriatric Depression Scale.

As indicated previously, participants were categorized – using the NIA-AA Research Framework based on PET-imaging of β-amyloid (A) and tau (T) deposition – as Normal AD Biomarkers (A-T-; *n* = 242), Alzheimer's Pathologic Change (A + T-; *n* = 116), or Alzheimer's disease (A + T+; *n* = 83). Diagnostic composition of participants from each biomarker status group after applying ADNI3 criteria (for Logical Memory, MMSE, and CDR

Table 1. Demographic, neuropsychological, and behavioral variables for the biomarker status groups and total sample.

Variable	A-T-	A ± T-	A ± T±	Total Sample
n	242	116	83	441
Age (years) [1]	69.96 (6.8)	71.82 (7.4)	72.65 (7.1)	70.95 (7.1)
Education (years)	16.57 (2.3)	16.41 (2.4)	16.23 (2.5)	16.46 (2.3)
Sex (% female)	57.9%	56.9%	50.6%	56.2%
Race (% Caucasian)	84.3%	88.8%	83.1%	85.3%
Geriatric Depression Scale	1.03 (1.3)	1.03 (1.2)	1.46 (1.6)	1.11 (1.4)
AMNART Verbal Intelligence	119.38 (9.6)	117.77 (10.3)	117.15 (9.9)	118.52 (9.9)
Baseline Diagnosis	NL = 172 MCI = 62 AD = 8	NL = 70 MCI = 35 AD = 11	NL = 15 MCI = 40 AD = 28	NL = 257 MCI = 137 AD = 47
MOCA [2,3]	25.46 (3.1)	24.38 (4.3)	20.43 (5.1)	24.23 (4.3)
RAVLT Total Recall [2, 3]	44.78 (10.9)	41.34 (12.4)	30.63 (12.2)	41.21 (12.7)
RAVLT Delayed Recall [2, 3, 4]	7.58 (4.8)	5.86 (4.3)	3.10 (5.1)	6.29 (5.0)
Logical Memory Immediate Recall [2, 3, 4]	12.94 (4.1)	11.43 (4.9)	7.86 (4.9)	11.58 (4.8)
Logical Memory Delayed Recall [2, 3]	11.45 (4.5)	10.05 (5.2)	5.57(5.2)	9.97 (5.3)
CDR-SB [2, 3]	0.57 (1.2)	0.91 (1.5)	2.18 (1.9)	0.96 (1.6)

Notes: AMNART Verbal Intellect score listed as a *Standard Score*, and all other scores are raw scores. All values are *Mean* (*Standard Deviation*) unless listed otherwise. [1] Denotes significant difference between A-T- and A + T+ groups, $p < .01$. [2] Denotes significant difference between A-T- and A + T+ groups, $p < .001$. **3** Denotes significant difference between A + T- and A + T+ groups, $p < .001$. [4] Denotes significant difference between A-T- and A + T- groups, $p < .001$.

Abbreviations: A-T- = Normal AD Biomarkers, A + T- = Alzheimer's Pathologic Change, A + T+ = Alzheimer's disease, AMNART = American National Adult Reading Test, NL = Normal Cognition, MCI = Mild Cognitive Impairment, AD = Alzheimer's Disease, MOCA = Montreal Cognitive Assessment, RAVLT = Rey Auditory Verbal Learning Test, CDR-SB = Clinical Dementia Rating Scale – Sum of Boxes.

performance) included the following phenotypes: 71.1% of the A-T- group was classified as having normal cognition, 25.6% as having MCI, and 3.3% as having AD; 60.3% of the A + T- group was classified as having normal cognition, 30.2% as having MCI, and 9.5% as having AD; and 48.2% of the A + T+ group was classified as having MCI, 33.7% as having AD, and 18.1% as having normal cognition. When comparing demographic differences across biomarker status groups, Table 1 shows that the A + T+ group possessed significantly greater proportions of participants diagnosed with AD than the other groups, $X^2(4) = 93.98$, $p < .001$, *Phi* = .46. A significant difference was observed between biomarker status groups for age, $F(2, 436) = 5.67$, $p = .004$, $\eta^2 = 0.03$, such that the A + T + group was older than the A-T- group ($p = .009$). No differences were observed for education, $F(2, 438) = 0.69$, $p = .50$, $\eta^2 = 0.003$, verbal intellect, $F(2, 424) = 2.00$, $p = .14$, $\eta^2 = 0.01$, self-reported depression, $F(2, 437) = 3.21$, $p = .05$, $\eta^2 = 0.01$, sex, $X^2(2) = 1.35$, $p = .51$, *Phi* = .06, or ethnicity, $X^2(2) = 1.63$, $p = .44$, *Phi* = .06. The A + T+ group performed significantly worse on all immediate and delayed memory measures (Total/Immediate and Delayed memory indexes from both RAVLT and Logical Memory) and global measures (MOCA and CDR-Sum of Boxes scores) than A + T- group (all *p*s < .001), who performed worse than the A-T- group (all *p*s < .001). Expanded results (e.g., *F* – values, effect sizes) are available upon request.

The mean value for LR derived from the ADAS-Cog Word Recall subtest in the total sample was 0.53 (*SD* = 0.3). This equates to the sample, on average, learning 53% of the available information after Trial 1. The mean value for RLS was 2.52 (*SD* = 1.6), and the mean value for LOT was 4.58 (*SD* = 2.6). See, Table 2. The bivariate correlation coefficient between LR and age was significant, $r = -.17$, $R^2 = .03$, $p < .001$, as were the correlation coefficients between LR and education, $r = .11$, $R^2 = .01$, $p = .01$, verbal intellect, $r = .16$,

Table 2. ADAS-Cog memory subtest and learning slope scores for the biomarker groups and total sample.

Variable	A-T-	A ± T-	A ± T±	Total Sample
n	242	116	83	441
Word Recall Immediate Learning [1, 2, 3]	20.93 (4.2; 3.00–30.00)	19.00 (5.1; 5.00–30.00)	14.73 (5.2; 3.00–25.00)	19.26 (5.2; 3.00–30.00)
Word Recall Delayed [1, 2, 3]	7.04 (2.2; 1.00–10.00)	6.22 (2.4; 2.00–10.00)	3.73 (2.4; 1.00–9.00)	6.20 (2.6; 1.00–10.00)
ADAS-Cog Total Score [1, 2, 3]	13.99 (5.6; 4.00–37.33)	16.41 (7.1; 5.33–38.33)	25.03 (9.4; 7.67–48.33)	16.71 (8.0; 4.00–48.33)
Word Recall LR [1, 2]	0.59 (0.4; -2.33–1.00)	0.50 (0.3; -0.50–1.00)	0.36 (0.3; -0.33–1.00)	0.53 (0.3; -2.33–1.00)
Word Recall RLS	2.64 (1.7; -7.00–6.00)	2.48 (1.6; -3.00–7.00)	2.24 (1.5; -2.00–5.00)	2.52 (1.6; -7.00–7.00)
Word Recall LOT	4.82 (2.6; -5.00–11.00)	4.48 (2.8; -4.00–14.00)	4.04 (2.5; -4.00–9.00)	4.58 (2.6; -5.00–14.00)

Notes: [1]Denotes significant difference between A-T- and A + T+ groups, $p < .001$. **2** Denotes significant difference between A + T- and A + T+ groups, $p < .001$. [3] Denotes significant difference between A-T- and A + T- groups, $p < .001$.
Abbreviations: ADAS-Cog = Alzheimer's Disease Assessment Scale – Cognitive Subscale, A-T- = Normal AD Biomarkers, A + T- = Alzheimer's Pathologic Change, A + T+ = Alzheimer's disease, LR = Learning Ratio, RLS = Raw Learning Score, LOT = Learning Over Trials. All values are *Mean* (*Standard Deviation; range*) unless listed otherwise.

$R^2 = .03$, $p < .001$, and sex, $r = .17$, $R^2 = .03$, $p < .001$. Similarly, bivariate correlations between age and both ADAS-Cog Word Recall Immediate and Delayed Recall were significant, $r = -.21$ to $-.25$, $R^2 = .04$ to .06, $ps < .001$, as were the correlation coefficients between Word Recall subtests and education, $rs = .19$ to .21, $R^2 = .04$, $ps < .001$, verbal intellect, $rs = .25$ to .33, $R^2 = .06$ to .11, $ps < .001$, and sex, $rs = .22$ to .27, $R^2 = .05$ to .07, $ps < .001$. No relationship between ethnicity and LR, Word Recall Immediate Recall, or Word Recall Delayed Recall was observed, $rs = .008$ to .06, $R^2 = .000$ to .003, $ps = .22$ to .87. Consequently, age, education, verbal intellect, and sex were used as covariates in the subsequent comparisons.

Criterion Validity Analyses and Test Operating Characteristics

When comparing ADAS-Cog memory subtests, learning slopes, and Total Score performances between individuals across A-T-, A + T-, and A + T+ groups, statistically significant differences were observed, *Wilk's Lamba* = .730, $F(12, 820) = 11.65$, $p < .001$, $\eta^2 = .15$, after controlling for age, education, verbal intellect, and sex. Specifically, group differences existed for Word Recall Immediate Recall, $F(2, 415) = 48.95$, $p < .001$, $\eta^2 = .19$, Word Recall Delayed Recall, $F(2, 415) = 56.78$, $p < .001$, $\eta^2 = .22$, Word Recall LR, $F(2, 415) = 9.59$, $p < .001$, $\eta^2 = .05$, and ADAS-Cog Total Score, $F(2, 415) = 72.71$, $p < .001$, $\eta^2 = .26$. As can be seen from the means in Table 2, the A + T+ group performed significantly worse than the A + T- and A-T- groups on all four variables ($ps < .001$), and the A + T- group performed worse than the A-T- group for Word Recall Immediate Recall and Delayed Recall (but not ADAS-Cog Total Score or Word Recall LR). Conversely, no differences were observed between biomarker status groups for the RLS, $F(2, 415) = 1.31$, $p = .27$, $\eta^2 = .006$, or LOT, $F(2, 415) = 2.53$, $p = .08$, $\eta^2 = .01$.

Table 3 displays the ROC-AUC values for the ADAS-Cog memory subtests, learning slopes, and Total Score when differentiating individuals between the A + T+ and A-T- groups. Each of the ADAS-Cog Word Recall Immediate Recall, ADAS-Cog Word Recall Delayed Recall, and ADAS-Cog Total Score AUC values possessed good AUC values, which

Table 3. Receiver operating characteristic area under curve, cut scores, and sensitivity/specificity when differentiating A-T- from A + T+ biomarker groups for ADAS-Cog Total Score, memory subtests, and learning slopes across the total sample.

Variable	AUC	95% CI	Cut score	Sensitivity/ Specificity
Word Recall Immediate Recall	0.820	0.765–0.875	≤17.50	0.723/0.822
Word Recall Delayed Recall	0.828	0.774–0.882	≤3.50	0.627/0.921
ADAS-Cog Total Score	0.828	0.774–0.882	≥15.835 *	0.807/0.725
Word Recall LR	0.746	0.688–0.805	≤0.5857	0.843/0.574
Word Recall RLS	0.574	0.503–0.644	≤3.50	0.795/0.289
Word Recall LOT	0.577	0.507–0.648	≤7.50	0.952/0.153

Notes: *Denotes the use of the test developer's original scoring procedures, where higher scores mean worse performance.

Abbreviations: ADAS-Cog = Alzheimer's Disease Assessment Scale – Cognitive Subscale, A-T- = Normal AD Biomarkers, A + T+ = Alzheimer's disease, AUC = Area Under Curve, CI = Confidence Interval, LR = Learning Ratio, RLS = Raw Learning Score, LOT = Learning Over Trials.

ranged from 0.820 to 0.828 (95% *CIs* ranging from 0.765 to 0.882). Additionally, the Word Recall LR metric possessed a fair AUC value (0.746; 95% *CI* of 0.688–0.805), whereas the RLS and LOT both failed to discern between biomarker groups (AUCs = 0.574 to 0.577; 95% *CIs* ranging from 0.503 to 0.648). While the values for the memory subtests and Total Score trended toward being larger than the AUC value for the LR metric, all the 95% CIs overlapped, suggesting no significant differences in the AUC values for LR versus the former measures (see, Table 3). Conversely, when comparing LR versus RLS and LOT AUC values, the AUCs were significantly larger for the LR metric, based on 95% CIs failing to overlap.

Also observed in Table 3, we derived cut scores for the ADAS-Cog Word Recall memory subtests, learning slope metrics, and ADAS-Cog Total Score to produce the highest balance of sensitivity and specificity. A cut score of ≤17.50 for Word Recall Immediate had a sensitivity of 0.723 and a specificity of 0.822. A cut score of ≤3.50 for Word Recall Delayed had a sensitivity of 0.627, but a specificity of 0.921, and a cut score of ≥15.835 for Total score had a sensitivity of 0.807 and a specificity of 0.725. Note that this latter score is "≥" because of the reverse scoring of the ADAS-Cog Total Score. When considering learning slope metrics, the sensitivity and specificity data similarly tended to be stronger for the LR metrics than either RLS or LOT metrics. For example, a cut score of ≤0.5857 for LR had sensitivity of 0.843 and specificity of 0.574, whereas a cut score of ≤4.0 for RLS had sensitivity of 0.795 and specificity of 0.289, and a cut score of ≤7.5 for LOT had sensitivity of 0.952 and a specificity of 0.153.

Convergent Validity Analyses

As indicated previously, the differences observed between learning slopes in the above analyses led us to consider supplementary convergent validity analyses between LR, RLS, and LOT. After accounting for age, education, verbal intellect, and sex, in the total sample Word Recall LR was significantly and positively related to immediate and delayed memory performances for not only the ADAS-Cog Word Recall subtest, but also for additional verbal memory measures (RAVLT and Logical Memory; all *ps* < .001; see, Table 4). Similarly, LOT was significantly and positively related to most immediate and delayed memory performances, but RLS was not significantly related to either memory measure from which it was not derived. When comparing across learning slopes, LR score correlations were

Table 4. Partial correlation coefficients between ADAS-Cog Word Recall learning slope scores and immediate/delayed recall scores across biomarkers status groups and the total sample, after controlling for age, education, verbal intellect, and sex.

Variable	Correlated with	A-T-	A ± T-	A ± T±	Total Sample
n		242	116	83	441
Word Recall LR	RAVLT Total Recall	.26, $p < .001$	.52, $p < .001$	.42, $p < .001$	.40, $p < .001$
	RAVLT Delayed Recall	.08, $p = .23$	.46, $p < .001$	.25, $p = .03$	.25, $p < .001$
	Logical Memory Immediate Recall	.14, $p = .04$	.40, $p < .001$	.42, $p < .001$	.31, $p < .001$
	Logical Memory Delayed	.13, $p = .06$	.42, $p < .001$	.43, $p < .001$	.30, $p < .001$
	Word Recall Immediate Recall	.46, $p < .001$	.57, $p < .001$	.63, $p < .001$	.53, $p < .001$
	Word Recall Delayed Recall	.26, $p < .001$	.55, $p < .001$	.57, $p < .001$	.43, $p < .001$
Word Recall RLS	RAVLT Total Recall	−.001, $p = .99$	.24, $p = .01$	.21, $p = .07$	.12, $p = .02$
	RAVLT Delayed Recall	−.07, $p = .28$	.19, $p = .05$	.15, $p = .18$	.05, $p = .32$
	Logical Memory Immediate Recall	.01, $p = .87$	.20, $p = .03$	.26, $p = .03$	.12, $p = .02$
	Logical Memory Delayed	−.01, $p = .86$	.26, $p = .007$	.25, $p = .03$	.12, $p = .02$
	Word Recall Immediate Recall	−.02, $p = .75$	.24, $p = .01$	.32, $p = .004$	.13, $p = .007$
	Word Recall Delayed Recall	−.009, $p = .89$	.31, $p < .001$	.31, $p = .006$	.15, $p = .002$
Word Recall LOT	RAVLT Total Recall	.06, $p = .35$	.29, $p = .002$	.26, $p = .02$	.19, $p < .001$
	RAVLT Delayed Recall	−.02, $p = .73$	.21 $p = .03$	.15, $p = .20$	.10, $p = .05$
	Logical Memory Immediate Recall	.07, $p = .28$	.28, $p = .003$	.29, $p = .01$	.20, $p < .001$
	Logical Memory Delayed	.04, $p = .57$	.35, $p < .001$	.32, $p = .005$	.20 $p < .001$
	Word Recall Immediate Recall	−.04, $p = .60$	.29, $p = .002$	.41, $p < .001$	.18, $p < .001$
	Word Recall Delayed Recall	.05, $p = .46$	.37, $p < .001$	.29, $p = .01$	.21, $p < .001$

Abbreviations: ADAS-Cog = Alzheimer's Disease Assessment Scale – Cognitive Subscale, A-T- = Normal AD Biomarkers, A + T- = Alzheimer's Pathologic Change, A + T+ = Alzheimer's disease, LR = Learning Ratio, RAVLT = Rey Auditory Verbal Learning Test, RLS = Raw Learning Score, LOT = Learning Over Trials.

consistently larger than RLS and LOT score correlations. Specifically, Fisher *r* to *z* transformations indicated that partial correlations were significantly greater for LR than RLS for all comparisons (*zs* = 2.75 to 6.66, *ps* .001 to .006). Partial correlations for LR were larger than LOT for all comparisons, reaching statistical significance for RAVLT Total Recall ($z = 3.26$, $p = .001$), ADAS-Cog Word Recall Immediate Recall ($z = 5.99$, $p < .001$), and ADAS-Cog Word Recall Delayed Recall ($z = 3.18$, $p = .002$). As can also be observed in Table 4, these effects were similarly observed when examining partial correlations individually within the A-T-, A + T-, and A + T+ groups.

Discussion

The current study reflects the first attempt to characterize the sensitivity of memory subtests and learning slopes from the ADAS-Cog at detecting differences in AD biomarker status, when applying NIA-AA Research Framework based on amyloid-PET and tau-PET imaging to the ADNI sample. Our results in Table 3 indicate that not only was the ADAS-Cog Total Score effective at discriminating participants with Normal AD Biomarkers (A-T-) from those with Alzheimer's Pathologic Change (A + T-) and AD (A + T+), the Word Recall Immediate Recall and Delayed Recall subtests (Question 1: Immediate Recall and Question 4: Delayed Recall) from the ADAS-Cog and the LR learning slope metric were effective as well. Specifically, the ADAS-Cog Total Score, ADAS-Cog Word Recall Immediate, and ADAS-Cog Word Recall Delayed subtests were all categorized as having "good" sensitivity at classifying biomarker status, and the LR learning slope was categorized as being "fair." While the Total Score result is not necessarily surprising given the multitude of studies relating ADAS-Cog Total Score to AD biomarkers (Avgerinos et al., 2021; Devous et al., 2021; Li et al., 2016), the latter results lend support for the use of ADAS-Cog memory

subtests in identifying amyloid and tau status. When examining the effect sizes between biomarker status groups and the Word Recall subtests (η^2s = .19 and .22) versus the commonly administered RAVLT Immediate and Delayed Recall subtests (η^2s = .17 and .10), the 10-item 3-trial Word Recall task performed comparably to – if not slightly better than – the 15-item 5-trial RAVLT. Beyond this statistical significance, the magnitude of the difference in ADAS-Cog Word Recall Immediate and Delayed Recall subtests between AD biomarker groups appears to be clinically meaningful. As seen in Table 2, A + T+ participants performed approximately 1.5 *SD* below their A-T- counterparts on both Immediate Recall and Delayed Recall subtests. This is also similar to the degree of clinical significance seen between amyloid-positive and amyloid-negative groups assessed using ^{18}F-Flutemetamol (Hammers et al., 2017) on a comparable word-list learning task (10-items over 4 trials; 1.7 *SD* to 2.3 *SD*) from the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 2012). These Word Recall results also coincide with findings of previous studies using a variety of other PET radioligands and word-list tasks that showed that greater amyloid and tau biomarker imaging is associated with worse memory performance (Ciampa et al., 2022; Farrell et al., 2022; Kawas et al., 2013; Lagarde et al., 2022; Ossenkoppele et al., 2014; Sperling et al., 2013), and provide further support for the typical clinical manifestation of AD beginning with early and predominant anterograde episodic memory deficits and being associated with amyloid plaque and tau accumulation.

The results of the current study similarly support the use of LR learning slope in discriminating amyloid and tau status. Although the AUC values for LR were slightly smaller than those of either the ADAS-Cog Total score or the ADAS-Cog memory subtests, these differences were not statistically significant and the LR AUC values were beyond cutoffs for acceptability (Hosmer et al., 2013). These results are consistent with the limited research in the literature examining learning slopes and AD biomarkers of amyloid. Thomas and colleagues (Thomas et al., 2020) successfully used LOT scores from the RAVLT to predict future amyloid accumulation in participants with objective subtle cognitive difficulties, and LR performance has previously been shown to be associated with amyloid deposition using ^{18}F-Flutemetamol (Hammers, Suhrie, Dixon, Gradwohl, Archibald et al., 2021). To date, our finding appears to represent the first association between learning slopes and tau pathology. Additionally, cut scores are displayed in Table 3 for ADAS-Cog Word Recall subtests, ADAS-Cog Total Score, and learning slopes to produce the highest balance of sensitivity and specificity. Similar to the AUC values, the sensitivity and specificity data tended to be slightly larger for ADAS-Cog memory subtests and Total Score than the LR metric, though all were generally comparable. For example, a cut score of ≤0.5857 for LR had sensitivity of 0.843 and specificity of 0.574, with cut scores for ADAS-Cog Word Recall Immediate Recall, ADAS-Cog Word Recall Delayed Recall, and ADAS-Cog Total Score all reflecting trade-offs between sensitivity and specificity relative to LR. For LR specifically, these results are consistent with previous research showing that the same cut score for LR (≤0.5857) had a sensitivity of 0.873 and a specificity of 0.846 when differentiating cognitive impaired and intact older adults when derived from the aforementioned word-list learning subtest from the RBANS (Hammers, Suhrie, Dixon, Gradwohl, Duff et al., 2021).

At a more basic level, these findings additionally provide both criterion and convergent validation for the development of a learning slope metric from the Word Recall subtest of

the ADAS-Cog. While the LR calculation has been applied to other word-list tasks in the literature (Hammers et al., 2022; Hammers, Suhrie et al., 2021; Spencer et al., 2020), the use of this process score for the Word Recall subtest has been previously unexamined and unvalidated. The aforementioned ability for LR to discriminate between amyloid and tau status provides evidence of criterion validity. Additionally, after accounting for age, education, verbal intellect, and sex, lower LR performance was associated with worse performance on learning and memory tests. Table 4 indicates that LR performance from the Word Recall subtest was positively and significantly related to learning and memory from both tasks from which LR was derived (Word Recall subtests) and other memory tasks from which LR was not derived (e.g., RAVLT and Logical Memory). Relative to research with LR metrics derived from other word-list measures (HVLT-R, RBANS List Learning, RAVLT; Hammers et al., 2022; Hammers, Suhrie et al., 2021), the magnitude of correlations between LR and both other word-list-learning tasks (*r*s = .41 to .50) and other story memory tasks (*r*s = .26 to .38) were also comparable with our current findings (*r*s = .40 and .31, respectively). This comparability is particularly interesting given that some of these other word-list-learning tasks reflect semantic clustering of stimuli (e.g., pieces of furniture, dwellings; HVLT-R), which has been suggested to improve learning and recall (Manning & Kahana, 2012). Additionally, the ADAS-Cog Word List includes 10 words presented over 3 trials (for a total of 30 possible points), and the others range from 10 to 15 words across 3–5 trials (for a range of 36 to 75 possible points), yet results were similar regardless of trial length/number. This represents convergent validity for LR derived from the ADAS-Cog Word Recall subtest, and further suggests that the original word-lists' trial length, number, and use of semantic clustering of stimuli has limited influence on the resultant process score.

Further, when comparing learning slopes, LR consistently outperformed the other metrics (RLS and LOT). For example, AUC-ROC values for RLS and LOT when differentiating between A-T- and A + T+ biomarker status groups were well-below cutoffs for acceptability (0.574 to 0.577; Table 3), and ANCOVA comparing RLS and LOT values between groups were non-significant. Overall, these results suggest that these learning metrics failed to distinguish AD biomarker status. Additionally, partial correlations for RLS and LOT with learning and memory measures were consistently and significantly smaller than those observed for LR (Table 4). These results coincide with several other studies comparing these learning slopes derived from the RBANS (Spencer et al., 2020), HVLT-R (Hammers et al., 2021), and RAVLT (Hammers et al., 2022), and provide further support for the superiority of LR over the other learning slopes. As such, our findings suggest that if considering a learning process score from the ADAS-Cog for future research, LR is a good choice.

The following represent limitations to our current study. First, these results are unique to the ADAS-Cog Word Recall subtest and may not necessarily generalize to all verbally presented word-list tasks (though our comparable convergent validity for LR suggests that some generalizability is likely evident) – nor to verbally presented paragraph memory or visual memory measures. Also, this study did not examine all possible learning slope calculations, but only LR from Spencer et al. (2020) and the other most commonly used slope metrics. Relatedly, we calculated LR by taking the difference between the first and final trial; while some research has used the difference between the first and best score among the final few trials (Hammers et al., 2022), our method is commonly conducted among other LR studies and was presently done for (1) ease of calculation and (2) because the use of only three

learning trials for the ADAS-Cog Word Recall subtest limits the likelihood of the final trial not being the best trial. Second, because we recruited our sample from ADNI, they tended to be highly educated and mostly Caucasian. Future research should consider replicating these findings in more heterogeneous samples to broaden generalizability. ADNI also applies rigorous exclusion criteria that tends to be typical of industry-sponsored clinical trials, therefore this cohort of individuals across the AD biomarker spectrum might not be representative of the general population. Further, ADNI-specific modifications have been made to the test administration of several measures used in the ADNI battery (e.g., Logical Memory only includes "Story A"), and many common ADNI measures have not been updated to reflect test advancement over time (e.g., Logical Memory from the WMS-R was developed in 1987, and newer versions have been developed since that time). Third, based on standard test administration instructions, the ADAS-Cog Word Recall Delayed subtest was only administered approximately 10 minutes after the completion of the three trials of the word list. While this period is shorter than other word-list-learning tasks (20–30 minutes for HVLT-R and RAVLT), our findings indicate that the subtest was still capable of measuring delayed recall. Fourth, in the current study, we calculated Immediate Recall and Delayed Recall from the ADAS-Cog Word Recall in a different fashion than what test developers recommend (for Questions 1 and 4 of the ADAS-Cog). This was done for ease of understanding and comparability with other memory measures. As a result, partial correlations between LR and ADAS-Cog Word Recall subtests were positive in our study, whereas they would have been negative (but equal in magnitude) if using the original scoring criteria to have higher scores reflect worse impairment (i.e., # of words not recalled; Rosen et al., 1984). Finally, it is important to emphasize that the Research Framework criteria only speaks to underlying pathology of AD, and not clinical presentation. As such, the participants with "Alzheimer's disease" (A + T+) in our sample presented with a range of memory abilities, and were clinically classified as cognitively normal (18%), MCI (48%), to dementia due to AD (33%). The use of the Research Framework in this study does not constitute an endorsement by the authors of using a pathology-only criteria for the diagnosis of AD, but instead we strongly support continued clinical utility of the use of neuropsychological testing in the clinical diagnosis of dementia due to AD.

Despite these limitations, the current study provides evidence of sensitivity for the use of both memory subtests and the LR-learning slope from the ADAS-Cog in distinguishing AD biomarker status. As the ADAS-Cog is typically only considered as a Total score, these results provide researchers additional tools – at no additional cost – to aid in identification of the presence of AD pathology. Further investigation of these ADAS-Cog memory subtests and process scores into criterion validity with clinical diagnoses and normative comparisons are encouraged in the future to enhance their application for clinical use.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern Californi; Department of Biotechnology, Ministry of Science and Technology.

References

ADNI1. (2004). *Alzheimer's Disease Neuroimaging Initiative: ADNI Procedures Manual*. http://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI_GeneralProceduresManual.pdf

ADNI2. (2011). *Alzheimer's Disease Neuroimaging Initiative: ADNI2 Procedures Manual*. https://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf

ADNI3. (2016). *Alzheimer's Disease Neuroimaging Initiative: ADNI3 Procedures Manual*. Retrieved July 2021 https://adni.loni.usc.edu/wp-content/uploads/2012/10/ADNI3-Procedures-Manual_v3.0_20170627.pdf

ADNIGO. (2009). *Alzheimer's Disease Neuroimaging Initiative: ADNI Grand Opportunities Procedures Manual*. http://adni.loni.usc.edu/wp-content/uploads/2008/07/ADNI_GO_Procedures_Manual_06102011.pdf

Avgerinos, K. I., Ferrucci, L., & Kapogiannis, D. (2021). Effects of monoclonal antibodies against amyloid-β on clinical and biomarker outcomes and adverse event risks: A systematic review and meta-analysis of phase III RCTs in Alzheimer's disease. *Ageing Res Rev*, *68*(1), 101339. https://doi.org/10.1016/j.arr.2021.101339

Bender, A. R., Brandmaier, A. M., Duzel, S., Keresztes, A., Pasternak, O., Lindenberger, U., & Kuhn, S. (2020). Hippocampal Subfields and Limbic White Matter Jointly Predict Learning Rate in Older Adults. *Cereb Cortex*, *30*(4), 2465–2477. https://doi.org/10.1093/cercor/bhz252

Benge, J. F., Balsis, S., Geraci, L., Massman, P. J., & Doody, R. S. (2009). How well do the ADAS-cog and its subscales measure cognitive dysfunction in Alzheimer's disease? *Dement Geriatr Cogn Disord*, *28*(1), 63–69. https://doi.org/10.1159/000230709

Birks, J. (2006). Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database Syst Rev(1)*, *2006* (1) , CD005593. https://doi.org/10.1002/14651858.Cd005593. Published 2006 Jan 25.

Brandt, J., & Benedict, R. (1997). *Hopkins Verbal Learning Test-Revised*. Psychological Assessment Resources, Inc.

Ciampa, C. J., Parent, J. H., Harrison, T. M., Fain, R. M., Betts, M. J., Maass, A., Winer, J. R., Baker, S. L., Janabi, M., Furman, D. J., D'Esposito, M., Jagust, W. J., & Berry, A. S. (2022). Associations among locus coeruleus catecholamines, tau pathology, and memory in aging. *Neuropsychopharmacology*, *47*(5), 1106–1113. https://doi.org/10.1038/s41386-022-01269-6

Devous, M. D.,sSr., Fleisher, A. S., Pontecorvo, M. J., Lu, M., Siderowf, A., Navitsky, M., Kennedy, I., Southekal, S., Harris, T. S., & Mintun, M. A. (2021). Relationships Between Cognition and

Neuropathological Tau in Alzheimer's Disease Assessed by 18F Flortaucipir PET. *J Alzheimers Dis*, *80*(3), 1091–1104. https://doi.org/10.3233/JAD-200808

Farrell, M. E., Papp, K. V., Buckley, R. F., Jacobs, H. I. L., Schultz, A. P., Properzi, M. J., Vannini, P., Hanseeuw, B. J., Rentz, D. M., Johnson, K. A., & Sperling, R. A. (2022). Association of Emerging β-Amyloid and Tau Pathology With Early Cognitive Changes in Clinically Normal Older Adults. *Neurology*, *98*(15), e1512–e1524. https://doi.org/10.1212/wnl.0000000000200137

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Gifford, K. A., Phillips, J. S., Samuels, L. R., Lane, E. M., Bell, S. P., Liu, D., Hohman, T. J., Romano, R. R., s3rd, Fritzsche, L. R., Lu, Z., & Jefferson, A. L., & Alzheimer's Disease Neuroimaging, I. (2015). Associations between Verbal Learning Slope and Neuroimaging Markers across the Cognitive Aging Spectrum. *J Int Neuropsychol Soc*, *21*(6), 455–467. https://doi.org/10.1017/S1355617715000430

Grober, E., & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *J Clin Exp Neuropsychol*, *13*(6), 933–949. https://doi.org/10.1080/01688639108405109

Hammers, D. B., Atkinson, T. J., Dalley, B. C. A., Suhrie, K. R., Horn, K. P., Rasmussen, K. M., Beardmore, B. E., Burrell, L. D., Duff, K., & Hoffman, J. M. (2017). Amyloid Positivity Using [18F]Flutemetamol-PET and Cognitive Deficits in Nondemented Community-Dwelling Older Adults. *Am J Alzheimers Dis Other Demen*, *32*(6), 320–328. https://doi.org/10.1177/1533317517698795

Hammers, D. B., Gradwohl, B. D., Kucera, A., Abildskov, T., Wilde, E. A., & Spencer, R. J. (2021). Preliminary validation of a measure for learning slope for the HVLT-R and BVMT-R in older adults. *Cognitive and Behavioral Neurology*, *34*(3), 170–181. https://doi.org/10.1097/WNN.0000000000000277

Hammers, D. B., Spencer, R. J., & Apostolova, L. G. (2022). Validation of and Demographically Adjusted Normative Data for the Learning Ratio Derived from the RAVLT in Robustly Intact Older Adults. *Arch Clin Neuropsychol*, *37*(5), 981–993. https://doi.org/10.1093/arclin/acac002

Hammers, D. B., Suhrie, K. R., Dixon, A., Gradwohl, B. D., Archibald, Z. G., King, J. B., Spencer, R. J., Duff, K., & Hoffman, J. M. (2021). Relationship between a novel learning slope metric and Alzheimer's disease biomarkers. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*, 6(1), 1–21. https://doi.org/10.1080/13825585.2021.1919984

Hammers, D. B., Suhrie, K., Dixon, A., Gradwohl, B. D., Duff, K., & Spencer, R. J. (2021). Validation of HVLT-R, BVMT-R, and RBANS Learning Slope Scores along the Alzheimer's Continuum. *Arch Clin Neuropsychol*. https://doi.org/10.1093/arclin/acab023

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression* (Third ed.). John Wiley & Sons, INC.

Kawas, C. H., Greenia, D. E., Bullain, S. S., Clark, C. M., Pontecorvo, M. J., Joshi, A. D., & Corrada, M. M. (2013). Amyloid imaging and cognitive decline in nondemented oldest-old: The 90+ Study. *Alzheimers Dement*, *9*(2), 199–203. https://doi.org/10.1016/j.jalz.2012.06.005

Kueper, J. K., Speechley, M., & Montero-Odasso, M. (2018). The Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): Modifications and Responsiveness in Pre-Dementia Populations. A Narrative Review. *J Alzheimers Dis*, *63*(2), 423–444. https://doi.org/10.3233/jad-170991

Li, J. Q., Tan, L., Wang, H. F., Tan, M. S., Tan, L., Xu, W., Zhao, Q. F., Wang, J., Jiang, T., & Yu, J. T. (2016). Risk factors for predicting progression from mild cognitive impairment to Alzheimer's disease: A systematic review and meta-analysis of cohort studies. *J Neurol Neurosurg Psychiatry*, *87*(5), 476–484. https://doi.org/10.1136/jnnp-2014-310095

Lowe, D. A., Balsis, S., Benge, J. F., & Doody, R. (2015). Adding delayed recall to the ADAS-cog improves measurement precision in mild Alzheimer's disease: Implications for predicting instrumental activities of daily living. *Psychological Assessment*, *27*(4), 1234–1240. https://doi.org/10.1037/pas0000133

Manning, J. R., & Kahana, M. J. (2012). Interpreting semantic clustering effects in free recall. *Memory*, *20*(5), 511–517. https://doi.org/10.1080/09658211.2012.683010

Mohs, R. C., Knopman, D., Petersen, R. C., Ferris, S. H., Ernesto, C., Grundman, M., Sano, M., Bieliauskas, L., Geldmacher, D., Clark, C., & Thal, L. J. (1997). Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the *Alzheimer*'s Disease Assessment Scale that broaden its scope. The *Alzheimer*'s Disease Cooperative Study. *Alzheimer Dis Assoc Disord*, *11*(Suppl 2), S13–21. https://doi.org/10.1097/00002093-199700112-00003

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, *43*(11), 2412–2414. https://doi.org/10.1212/WNL.43.11.2412-a

Morrison, R. L., Pei, H., Novak, G., Kaufer, D. I., Welsh-Bohmer, K. A., Ruhmel, S., & Narayan, V. A. (2018). A computerized, self-administered test of verbal episodic memory in elderly patients with mild cognitive impairment and healthy participants: A randomized, crossover, validation study. *Alzheimers Dement (Amst)*, *10*(1), 647–656. https://doi.org/10.1016/j.dadm.2018.08.010

Nasreddine, Z. S., Phillips, N. A., Bedirian, 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. *J Am Geriatr Soc*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Jack, C. R.,sJr., Bennett, D. A., Blennow, K. et al. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement*, *14*(4), 535–562. doi:10.1016/j.jalz.2018.02.018.

Ossenkoppele, R., Madison, C., Oh, H., Wirth, M., van Berckel, B. N., & Jagust, W. J. (2014). Is verbal episodic memory in elderly with amyloid deposits preserved through altered neuronal function? [Research Support, N.I.H., Extramural]. *Cereb Cortex*, *24*(8), 2210–2218. https://doi.org/10.1093/cercor/bht076

Randolph, C. (2012). *Repeatable Battery for the Assessment of Neuropsychological Status*. The Psychological Corporation.

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *Am J Psychiatry*, *141*(11), 1356–1364. https://doi.org/10.1176/ajp.141.11.1356

Sano, M., Raman, R., Emond, J., Thomas, R. G., Petersen, R., Schneider, L. S., & Aisen, P. S. (2011). Adding delayed recall to the *Alzheimer* Disease Assessment Scale is useful in studies of mild cognitive impairment but not *Alzheimer* disease. *Alzheimer Dis Assoc Disord*, *25*(2), 122–127. https://doi.org/10.1097/WAD.0b013e3181f883b7

Schmidt, M. (1996). *The Rey Auditory Verbal Learning Test*. Western Psychological Services.

Schwarz, A. J., Shcherbinin, S., Slieker, L. J., Risacher, S. L., Charil, A., Irizarry, M. C., Fleisher, A. S., Southekal, S., Joshi, A. D., Devous, M. D.,sSr., Miller, B. B., & Saykin, A. J. (2018). Topographic staging of tau positron emission tomography images. *Alzheimers Dement (Amst)*, *10*(1), 221–231. https://doi.org/10.1016/j.dadm.2018.01.006

Sheikh, J. I., & Yesavage, J. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist*, *5*(1–2), 165–172. https://doi.org/10.1300/J018v05n01_09

Spencer, R. J., Gradwohl, B. D., Williams, T. F., Kordovski, V. M., & Hammers, D. B. (2020). Developing learning slope scores for the repeatable battery for the assessment of neuropsychological status. *Appl Neuropsychol Adult*, *27*(1), 1–7. https://doi.org/10.1080/23279095.2020.1791870

Sperling, R. A., Johnson, K. A., Doraiswamy, P. M., Reiman, E. M., Fleisher, A. S., Sabbagh, M. N., Sadowsky, C. H., Carpenter, A., Davis, M. D., Lu, M., Flitter, M., Joshi, A. D., Clark, C. M., Grundman, M., Mintun, M. A., Skovronsky, D. M., & Pontecorvo, M. J. (2013). Amyloid deposition detected with florbetapir F 18 ((18)F-AV-45) is related to lower episodic memory performance in clinically normal older individuals. *Neurobiol Aging*, *34*(3), 822–831. https://doi.org/10.1016/j.neurobiolaging.2012.06.014

Lagarde, J., Olivieri, P., Tonietto, M. et al. (2022). Tau-PET imaging predicts cognitive decline and brain atrophy progression in early Alzheimer's disease. *J Neurol Neurosurg Psychiatry*, *93*(5), 459–467. doi:10.1136/jnnp-2021-328623 .

Thomas, K. R., Bangen, K. J., Weigand, A. J., Edmonds, E. C., Wong, C. G., Cooper, S., Delano-Wood, L., Bondi, M. W., & Alzheimer's Disease Neuroimaging, I. (2020). Objective subtle cognitive difficulties

predict future amyloid accumulation and neurodegeneration. *Neurology*, *94*(4), e397–e406. https://doi.org/10.1212/WNL.0000000000008838

Wechsler, D. (1987). *WMS-R: Wechsler Memory Scale–Revised: Manual*. The Psychological Corporation.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., sJr., Jagust, W., Morris, J. C., Petersen, R. C., Salazar, J., Saykin, A. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., & Alzheimer's Disease Neuroimaging, I. (2017). The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimers Dement*, *13*(5), 561–571. https://doi.org/10.1016/j.jalz.2016.10.006

Wesnes, K. A. (2008). Assessing Change in Cognitive Function in Dementia: The Relative Utilities of the Alzheimer's Disease Assessment Scale – Cognitive Subscale and the Cognitive Drug Research System. *Neurodegenerative Diseases*, *5*(3–4), 261–263. https://doi.org/10.1159/000113719

Wolk, D., Salloway, S., & Dickerson, B. (2019). Putting the New Alzheimer Disease Amyloid, Tau, Neurodegeneration (AT[N]) Diagnostic System to the Test. *JAMA*, *321*(23), 2289–2291. https://doi.org/10.1001/jama.2019.7534