



# Optimizing Cognitive Assessment Outcome Measures for Alzheimer's Disease by Matching Wordlist Memory Test Features to Scoring Methodology

Jason R. Bock\*, Julie Russell, Junko Hara and Dennis Fortier

*Embic Corporation, Newport Beach, CA, United States*

## OPEN ACCESS

### Edited by:

John Edward Harrison,  
Metis Cognition Ltd, United Kingdom

### Reviewed by:

Mike Conway,  
The University of Utah, United States  
Ivan Miguel Pires,  
Universidade da Beira  
Interior, Portugal

### \*Correspondence:

Jason R. Bock  
jrbock@embic.us

### Specialty section:

This article was submitted to  
Connected Health,  
a section of the journal  
Frontiers in Digital Health

**Received:** 30 July 2021

**Accepted:** 11 October 2021

**Published:** 03 November 2021

### Citation:

Bock JR, Russell J, Hara J and  
Fortier D (2021) Optimizing Cognitive  
Assessment Outcome Measures for  
Alzheimer's Disease by Matching  
Wordlist Memory Test Features to  
Scoring Methodology.  
*Front. Digit. Health* 3:750549.  
doi: 10.3389/fgdh.2021.750549

Cognitive assessment with wordlist memory tests is a cost-effective and non-invasive method of identifying cognitive changes due to Alzheimer's disease and measuring clinical outcomes. However, with a rising need for more precise and granular measures of cognitive changes, especially in earlier or preclinical stages of Alzheimer's disease, traditional scoring methods have failed to provide adequate accuracy and information. Well-validated and widely adopted wordlist memory tests vary in many ways, including list length, number of learning trials, order of word presentation across trials, and inclusion of semantic categories, and these differences meaningfully impact cognition. While many simple scoring methods fail to account for the information that these features provide, extensive effort has been made to develop scoring methodologies, including the use of latent models that enable capture of this information for preclinical differentiation and prediction of cognitive changes. In this perspective article, we discuss prominent wordlist memory tests in use, their features, how different scoring methods fail or successfully capture the information these features provide, and recommendations for emerging cognitive models that optimally account for wordlist memory test features. Matching the use of such scoring methods to wordlist memory tests with appropriate features is key to obtaining precise measurement of subtle cognitive changes.

**Keywords:** Alzheimer's disease, cognitive assessment, wordlist memory tests, cognitive modeling, review, digital cognitive biomarkers, clinical trials, latent cognitive processes

## INTRODUCTION

Wordlist memory (WLM) tests are the most common measures of verbal episodic memory used in clinical and research settings (1, 2). They are frequently used to screen individuals prior to neuroimaging or other assessments for cognitive impairment or dementia stages of Alzheimer's disease (AD) and to monitor progressive decline and treatment effects (3). AD research has recently shifted its focus from mild cognitive impairment (MCI) and moderate AD stages toward asymptomatic or preclinical AD stages, in which the cognitive changes may be very subtle and difficult to measure (4). This has prompted the research community to examine the WLM tests that they use and develop more sophisticated scoring to achieve the greatest precision of measurement (5). A wide variety of WLM tests are in use, and each of them has a distinct set of features (e.g., wordlist length, fixed vs. shuffled word-order across trials, inclusion of semantic categories) which impact the way that individuals learn and remember the words presented in them.

In patients at risk for AD, performance on a WLM test is characterized by poorer learning, more rapid forgetting, intrusion errors, and poorer recognition that reflect pathological changes in brain regions specific to memory (6–10). A WLM test's capacity to predict AD progression at the earliest stage of change can be evaluated by examining its construction, administration procedures, and scoring methods (11). All WLM tests present a list of words over a number of learning trials and subsequently ask the examinee to freely recall as many words as they can, assessing both working memory and short-term memory across immediate learning and free recall trials as well as short-term memory alone during one or more delayed free recall trial(s). Beyond this core component, WLM tests also have varying features that impact an individual's ability to encode the presented list words into memory and retrieve them (8, 9, 12, 13). These features include the number of words to learn, properties of the words (e.g., concreteness, length, frequency, context variability, valence, and arousal), the number of learning trials, whether the list is presented in a fixed or shuffled word-order across trials, whether the words belong to semantic categories or are unrelated words, the length of the delay between learning trials and delayed free recall trial(s), whether the measure includes cued recall trials, whether it includes recognition trials, and whether those recognition trials use the same words or different words from the recall trials.

Within an individual's sequence of responding to a WLM test are distinct response patterns that, when effectively analyzed, are capable of differentiating individuals in asymptomatic, or preclinical, stages of AD from cognitively normal individuals (14). To achieve this, researchers must move away from simple summary scores (number of words recalled on a trial or in a test) and even composite scores, and toward more sophisticated methods of scoring, such as modeling of latent variables (15, 16). These approaches have improved WLM tests' capacity to identify subtle cognitive changes compared to traditional scoring. How well any WLM test can characterize cognitive performance jointly depends on the features of the test that produce the performance as well as how that performance is analyzed.

To the authors' knowledge, however, no widely-used scoring methodologies systematically take these features into account. Therefore, we discuss common WLM tests and their features as well as scoring methods and recommendations for appropriately matching them together to optimize measurement precision.

## PREVALENT WORDLIST MEMORY TESTS AND THEIR FEATURES

Numerous WLM tests are in use today with varying wordlist features and respective applicability to memory performance. Some well-established and widely used WLM tests of verbal episodic memory include the California Verbal Learning Test-Second Edition [CVLT-II; (17)], the Rey Auditory-Verbal Learning Test or simply Auditory Verbal Learning Test [RAVLT or AVLT; (18, 19)], the Hopkins Verbal Learning Test-Revised [HVLTR; (20)], the International Shopping List Test [ISLT; (21)], the Alzheimer's Disease Assessment Scale-Cognitive

Subscale [ADAS-Cog, Word Recall; (22)], the Consortium to Establish a Registry for Alzheimer's Disease neuropsychological battery [CERAD, Word List Test; (23)], and the MCI Screen [MCIS; (24)]. The CVLT-II, AVLT, and HVLTR are more often used in clinical settings as part of a larger flexible neuropsychological battery to assess cognitive function, and the word recall parts of the ADAS-Cog and CERAD are subtests of a fixed battery specifically designed to be used for the assessment of AD progression.

### Wordlist Memory Test Features

Each test has key features that provide advantages or disadvantages compared to other tests depending on the needs of the study. The CVLT-II and the AVLT are the longest of the word list tests, with the greatest number of words to learn and total trials, which can provide finer quantitative assessment of memory performance and greater sensitivity in distinguishing less impaired individuals and subtypes of memory impairment (11, 17, 19). However, the length of time to administer these tests and the cognitive demand required makes them less practical for research purposes and for assessing more impaired individuals (25, 26). The ADAS-Cog and the CERAD WLM subtests are shorter and have fewer trials, and due to their specific development for use in the assessment of AD, these are widely used in AD clinical research trials. The ISLT is also frequently used in AD clinical research trials; however, due to the use of words belonging to a single semantic category (food items found in a grocery store), proactive interference, a potentially useful marker of impairment in early AD, is reduced compared to tests with zero or more than one semantic category (27). The MCIS, adapted from the CERAD WLM test, includes additional feature-equivalent wordlists and uses computerized administration protocol and scoring software (24). **Table 1** summarizes commonly used WLM tests of episodic memory.

### Wordlist Outcome Measures

Depending on which is used, WLM tests may provide the following outcome measures: individual item response data (i.e., specific words recalled), levels of total recall and recognition (i.e., summary scores), learning strategy use (e.g., semantic clustering, serial clustering, subjective clustering), primacy and recency effects, rate of new learning or acquisition, consistency of item recall, degree of vulnerability to proactive and retroactive interference, retention/forgetting over short and longer delays, cueing and recognition performance, discriminability and response bias, analysis of intrusion-error types, repetition errors, and analysis of false-positive types in recognition testing (28, 29).

### Wordlist Measures of Recognition

The addition of a recognition trial on some WLM tests helps to differentiate individuals with suspected retrieval problems, who may score better on a recognition trial than on a delayed free recall trial (28). Patients with AD have reduced benefit from cueing on a WLM recognition task due to impaired ability to consolidate learned words (8). In addition, Clark et al. (30) found that patients with amnesic MCI had poorer recognition memory abilities on a WLM recognition

**TABLE 1** | Commonly used wordlist memory tests and their features.

Test name	# of words	Semantic categories	# of learning trials	Word order	List B	# of delayed trials	Time of longest delay	Cued trials	Old/new recognition trial
CVLT-II	16	4	5	Fixed	Yes	2	20 min	Yes	Yes
AVLT	15	No	5	Fixed	Yes	2	20 min	No	Yes
HLVT-R	12	3	3	Fixed	No	1	20–25 min	No	Yes
ISLT	12	1	3	Fixed	No	1	10–15 min	No	No
ADAS-Cog Word Recall	10	No	3	Shuffled	No	1	5–8 min	No	No
CERAD Word List	10	No	3	Shuffled	No	1	10 min	No	Yes
MCIS	10	No	3	Fixed	No	1	5 min	No	Yes

task compared to healthy controls, specifically driven by an increase in false-positive errors rather than a reduced number of correct responses. Their findings suggested that individuals with amnesic MCI are more sensitive to proactive interference than cognitively normal older adults. In addition, healthy older adults who took a WLM test and later developed MCI exhibited rapid decay of words 8 years prior to diagnosis, with worse recognition discriminability and a greater number of intrusion errors evident 2 years prior to diagnosis (10). In another study, investigating the relationship between WLM test performance and brain activity in cognitively normal individuals with the apolipoprotein E4 allele (a genetic risk factor for AD), compared to those without, Matura et al. (31) found that individuals with the E4 genotype showed comparatively impaired verbal recognition and cued recall memory on WLM tests. They also found a different resting state in the brain connectivity pattern between E4 carriers and non-carriers, with positive correlations between recognition discriminability scores and resting-state values in the left hemisphere of the brain associated with verbal episodic memory, suggesting a possible compensatory process occurring in this region. These study findings highlight the importance of quantifying cognitive processes, such as recognition discriminability, on WLM tests that feature a recognition component, in order to identify patterns of performance that indicate the presence of AD.

### Wordlist Measures of Serial Position

Examining serial position recall accuracy on WLM tests can also provide important information about individual differences in episodic verbal memory performance that can reveal deficits in memory encoding. The serial position effect, in which more words are recalled at the beginning (primacy) and end (recency) of a list than in the middle, is frequently analyzed in AD research (32). Patients with very mild to moderate AD exhibit a reduced primacy effect and a normal or increased recency effect (13, 33–35). Bruno et al. (36, 37) found that the ratio between immediate and delayed performance scores at the end of the list, is a sensitive marker of early MCI, with higher ratios suggesting greater risk for neurodegenerative pathology. Additionally, Tomadesso et al. (38) evaluated serial position effects in individuals with MCI who were positive for  $\beta$ -amyloid, a biomarker of AD, compared them to  $\beta$ -amyloid negative

groups, and found that the  $\beta$ -amyloid positive group exhibited worse primacy performance. A WLM measure's presentation of words in a fixed or shuffled order across learning trials will impact the serial position effect and its capacity to inform analyses. Fixed word-order presentation maintains and reinforces serial position effects across WLM test trials, while shuffled word-order presentation eliminates the per-trial serial position effects across trials.

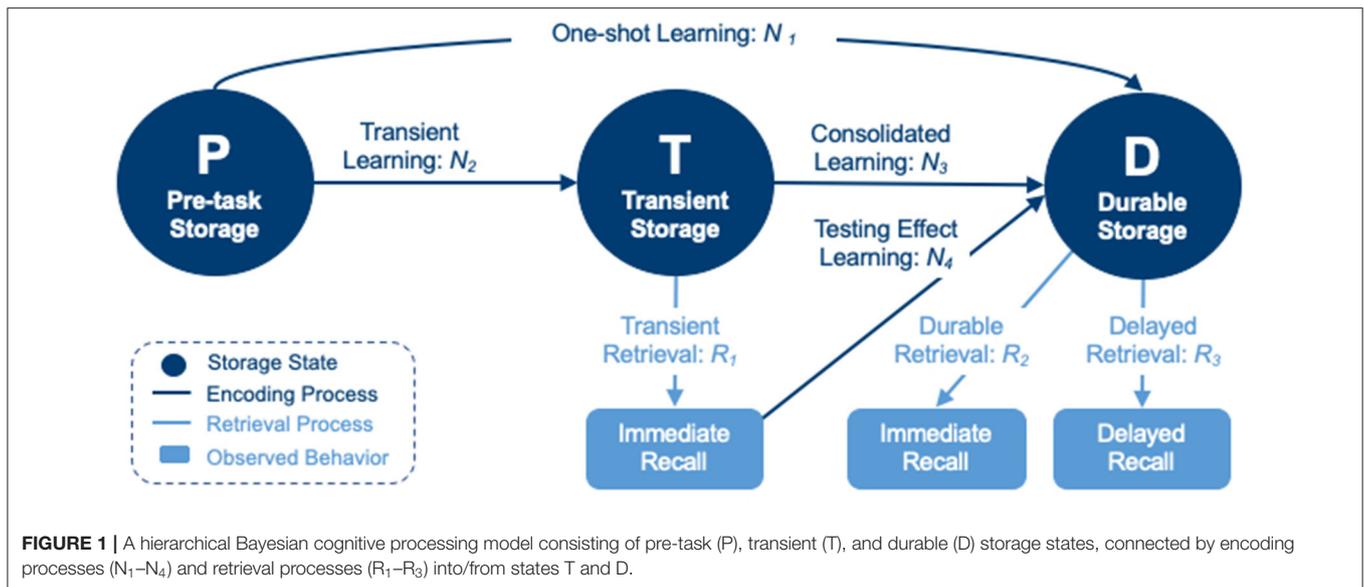
### ANALYTIC APPROACHES

While the literature examining WLM cognitive processes shows the meaningfulness of these more sensitive measures of performance in detecting and predicting underlying memory deficits, many AD studies and clinical trials continue to use summary or memory composite scores with a set cutoff that may be disproportionately impacted by poor performance in one area (39–41). This approach dilutes a specific impairment or treatment response and leads to inefficiencies throughout a clinical trial, from screening failures to response failures that may lead to premature discontinuation of a valuable treatment, as was seen in recent AD clinical trials (16, 42).

### Composite Scoring Approaches

To overcome the limitations of summary scores in assessing early or preclinical AD, researchers developed composite scores that combine information from across multiple WLM and other tests (5). An early composite score, the ADNI-Mem, incorporates several tests used in the longitudinal Alzheimer's Disease Neuroimaging Initiative (ADNI), including the AVLT and ADAS-Cog WLM tests, and performed similarly or "slightly better" than its constituent tests (43). Wang et al. (44) developed ADCOMS, including the ADAS-Cog WLM test, and better measured clinical progression in AD and MCI than constituent tests. However, these and other composite scoring methods have not consistently demonstrated the ability to distinguish preclinical AD from normal individuals (16). More precise methods are required.

Some statistical modeling approaches go deep rather than wide. The multivariate approach uses various features of a given WLM test to analyze cognitive processes underlying learning and memory. Thomas et al. (45) used a Cox proportional hazards



model to examine the added predictive value of individual cognitive processing variables (i.e., intrusion errors, learning slope, proactive interference, and retroactive interference) on a WLM test that included an interference word list. They found that intrusion errors contributed unique value in predicting progression from cognitively normal to MCI within 5 years. Another scoring model was developed for use with the MCIS, using an approach based on correspondence analysis of item response data and demographic covariates (24). This method is able to differentiate cognitively normal individuals from those with MCI with 97.3–99% accuracy (46, 47). These approaches demonstrate the value in using item-specific data from tests with complex features.

## Latent Modeling Approach

Due to high screen failure rates for  $\beta$ -amyloid PET when using traditional WLM test cutoff composite scores, a practical and sensitive WLM measure should be combined with a complex processing model to provide the greatest predictive capabilities. In comparison to the composite scoring approach, the latent modeling approach uses data captured by various features of WLM tests to analyze cognitive processes underlying test performance. In a simulation study, Proust-Lima et al. (15) compared inferences made by these two approaches and found that composite score risk factor accuracy is significantly reduced when constituent tests are not highly reliable or when there is systematically missing data, common in studies. In those cases, they recommend latent models.

One such model uses multinomial processing trees and hierarchical Bayesian computational methods to quantify encoding and retrieval processes of learning in multi-trial WLM tests (48). Using this hierarchical Bayesian cognitive processing (HBCP) model (Figure 1) in a recent wordlist study, researchers were able to generate digital cognitive biomarkers (DCB) for various encoding and retrieval processes that cannot be directly

observed or measured (49). These DCBs demonstrated the ability to distinguish groups of individuals with impending cognitive decline from those who would remain cognitively normal (14).

This class of model characterizes latent processes of information encoding and retrieval by utilizing item response data directly, and by building the effects of specific features that impact learning and memory directly into the model. Lee et al. (49) performed a nested analysis of a model that compared estimation of DCBs for each independent word against a model that calculated word-level DCBs from hierarchical estimations of primacy and recency directly, quantifying these features for comparison between impaired and non-impaired patients. However, this specific model relies on data that comes from a fixed word-order WLM test, as shuffled word-order WLM tests fail to produce a serial position effect. Similarly, such a model that incorporates recognition item responses would be able to quantify individuals' discriminability, and simultaneously model it with account of other cognitive processing parameters, when a WLM test includes recognition task data.

## DISCUSSION

There is a great variety of ways to implement WLM tests as well as ways to score them. It is imperative for the study of AD, as well as memory research in general, that the lessons learned from evaluating these tests over recent decades are put into practice. Regardless of the test and the features therein, summary scores are insufficient for detection of the subtle cognitive differences in early or preclinical AD (42). Composite scores are more informative by virtue of adding the information of multiple summary scores together, but these do not take into account the unique benefits of individual WLM test features (16, 43, 44). Nevertheless, WLM tests which remove effects (e.g., shuffled word-order or control for semantic similarity) are best scored with methods that do not or cannot account for

those effects. This is because greater or lesser performance for specific words of a list will produce increased error variance in methods not accounting for them, while removing these effects removes the error. However, there is valuable information in these performance differences, when a scoring method is able to account for them (32, 36, 37, 49). Expanding the scope of the data obtained from WLM tests through the use of more comprehensive analyses, such as with the described HBCP model, can significantly improve the efficiency of large-scale dementia research studies and provide valuable information about the efficacy of treatments, when paired with WLM tests that contain information produced by complex processes. While this approach has a limitation of increased complexity for interpretation and explanation of outcome measures, requiring sophistication in presenting results in clinical trials and healthcare settings, it greatly improves precision and granularity of information, compared to traditional approaches. This can be compared to machine learning, another sophisticated approach, which offers the greatest predictive capability but with even greater limitation in terms of interpretability (50). In all cases, matching the appropriate analytical method to the type of wordlist features in a given test will extract the greatest amount of information about performance and best illuminate patterns that both characterize cognitive deficits and predict cognitive change.

## CONCLUSION

Wordlist memory tests are commonly used for cognitive assessment, particularly in Alzheimer's disease research and

screening. Commonly used tests employ a variety of inherent features, such as list length, number of learning trials, order of presentation across trials, and inclusion of semantic categories. Historically, scoring methods, such as summary scores and more recently composite scoring, have not effectively addressed differences among these features, nor have they accounted for the manner in which they may modify learning and memory during task performance. Recent developments in latent modeling have shown great potential for using specific task features to accurately quantify the underlying cognitive processes used in learning and memory. Therefore, it is beneficial to match the features of a wordlist memory test to the appropriate scoring method that accounts for those particular features. Doing so facilitates the most precise characterization of cognitive performance and optimizes the likelihood of quantifying subtle but significant cognitive changes.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

JB, JH, and DF conceived and coordinated the writing of this article. JR wrote the first draft and identified references. All authors equally contributed to editing, revision, and approval of this article.

## REFERENCES

1. Belleville S, Fouquet C, Hudon C, Zomahoun HTV, Croteau J. Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. *Neuropsychol. Rev.* (2017) 27:328–53. doi: 10.1007/s11065-017-9361-5
2. Salmon DP, Bondi MW. Neuropsychological assessment of dementia. *Annu Rev Psychol.* (2009) 60:257–82. doi: 10.1146/annurev.psych.57.102904.190024
3. Lezak MD, Howieson DB, Bigler ED, Tranel D. *Neuropsychological Assessment*. 5th ed. Oxford: Oxford University Press (2012).
4. Rafii MS, Aisen PS. Alzheimer's disease clinical trials: moving toward successful prevention. *CNS Drugs.* (2019) 33:99–106. doi: 10.1007/s40263-018-0598-1
5. Aisen PS, Cummings J, Jack CR, Morris JC, Sperling R, Frölich L, et al. On the path to 2025: understanding the Alzheimer's disease continuum. *Alzheimer's Res Therapy.* (2017) 9:60. doi: 10.1186/s13195-017-0283-5
6. Bondi MW, Monsch AU, Galasko D, Butters N, Salmon DP, Delis DC. Preclinical cognitive markers of dementia of the Alzheimer type. *Neuropsychology.* (1994) 8:374–84. doi: 10.1037/0894-4105.8.3.374
7. Bondi MW, Salmon DP, Galasko D, Thomas RG, Thal LJ. Neuropsychological function and apolipoprotein E genotype in the preclinical detection of Alzheimer's disease. *Psychol Aging.* (1999) 14:295–303. doi: 10.1037/0882-7974.14.2.295
8. De Simone MS, Perri R, Fadda L, Caltagirone C, Carlesimo GA. Predicting progression to Alzheimer's disease in subjects with amnesic mild cognitive impairment using performance on recall and recognition tests. *J Neurol.* (2019) 266:102–11. doi: 10.1007/s00415-018-9108-0
9. Libon DJ, Bondi MW, Price CC, Lamar M, Eppig J, Wambach DM, et al. Verbal serial list learning in mild cognitive impairment: a profile analysis of interference, forgetting, and errors. *J Int Neuropsychol Soc.* (2011) 17:905–14. doi: 10.1017/S1355617711000944
10. Mistridis P, Krumm S, Monsch AU, Berres M, Taylor KI. The 12 years preceding mild cognitive impairment due to Alzheimer's disease: the temporal emergence of cognitive decline. *J Alzheimer's Dis.* (2015) 48:1095–107. doi: 10.3233/JAD-150137
11. Cerami C, Dubois B, Boccardi M, Monsch AU, Demonet JF, Cappa SF. Clinical validity of delayed recall tests as a gateway biomarker for Alzheimer's disease in the context of a structured 5-phase development framework. *Neurobiol Aging.* (2017) 52:153–66. doi: 10.1016/j.neurobiolaging.2016.03.034
12. Duke Han S, Nguyen CP, Stricker NH, Nation DA. Detectable Neuropsychological differences in early preclinical Alzheimer's disease: a meta-analysis. *Neuropsychol Rev.* (2017) 27:305–25. doi: 10.1007/s11065-017-9366-0
13. Martin ME, Sasson Y, Crivelli L, Gerschovich ER, Campos JA, Calcagno ML, et al. Relevance of the serial position effect in the differential diagnosis of mild cognitive impairment, Alzheimer-type dementia, and normal ageing. *Neurologia.* (2013) 28:219–25. doi: 10.1016/j.nrleng.2012.04.014
14. Bock JR, Hara J, Fortier D, Lee MD, Petersen RC, Shankle WR. Application of digital cognitive biomarkers for Alzheimer's disease: identifying cognitive process changes and impending cognitive decline. *J Prev Alzheimers Dis.* (2020) 8:123–6. doi: 10.14283/jpad.2020.63
15. Proust-Lima C, Philipps V, Dartigues JF, Bennett DA, Glymour MM, Jacqmin-Gadda H, et al. Are latent variable models preferable to composite score approaches when assessing risk factors of change? Evaluation of type-I error and statistical power in longitudinal cognitive studies. *Stat Methods Med Res.* (2019) 28:1942–57. doi: 10.1177/0962280217739658

16. Schneider LS, Goldberg TE. Composite cognitive and functional measures for early stage Alzheimer's disease trials. *Alzheimers Dement.* (2020) 12:e12017. doi: 10.1002/dad2.12017
17. Delis DC, Kramer JH, Kaplan E, Ober BA. *California Verbal Learning Test*. 2nd ed. San Antonio, TX: Psychological Corporation (2000).
18. Rey A. L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie.* (1941) 28:215–85.
19. Schmidt M. *Rey Auditory-Verbal Learning Test*. Los Angeles: Western Psychological Services (1996).
20. Benedict RHB, Schretlen D, Groninger L, Brandt J. Hopkins verbal learning test—revised: normative data and analysis of inter-form and test-retest reliability. *Clin Neuropsychol.* (1998) 12:43–55. doi: 10.1076/clin.12.1.43.1726
21. Lim YY, Prang KH, Cysique L, Pietrzak RH, Snyder PJ, Maruff P. A method for cross-cultural adaptation of a verbal memory assessment. *Behav Res Methods.* (2009) 41:1190–200. doi: 10.3758/BRM.41.4.1190
22. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry.* (1984) 141:1356–64. doi: 10.1176/ajp.141.11.1356
23. Morris JC, Mohs RC, Rogers H, Fillenbaum G, Heyman A. Consortium to establish a registry for Alzheimer's disease (CERAD) clinical and neuropsychological assessment of Alzheimer's disease. *Psychopharmacol Bull.* (1988) 24:641–52.
24. Shankle WR, Romney AK, Hara J, Fortier D, Dick MB, Chen JM, et al. Methods to improve the detection of mild cognitive impairment. *Proc Natl Acad Sci USA.* (2005) 102:4919–24. doi: 10.1073/pnas.0501157102
25. Beck IR, Gagneux-Zurbriggen A, Berres M, Taylor KI, Monsch AU. Comparison of verbal episodic memory measures: consortium to Establish a Registry for Alzheimer's disease—Neuropsychological Assessment Battery (CERAD-NAB) versus California Verbal Learning Test (CVLT). *Arch Clin Neuropsychol.* (2012) 27:510–9. doi: 10.1093/arclin/acs056
26. Lacritz LH, Cullum CM, Weiner MF, Rosenberg RN. Comparison of the hopkins verbal learning test-revised to the california verbal learning test in Alzheimer's disease. *Appl Neuropsychol.* (2001) 8:180–4. doi: 10.1207/S15324826AN0803\_8
27. Rahimi-Golkhandan S, Maruff P, Darby D, Wilson P. Barriers to repeated assessment of verbal learning and memory: a comparison of international shopping list task and rey auditory verbal learning test on build-up of proactive interference. *Arch Clin Neuropsychol.* (2012) 27:790–5. doi: 10.1093/arclin/acs074
28. Fine EM, Delis DC, Wette SR, Jacobson MW, Hamilton JM, Peavy G, et al. Identifying the “source” of recognition memory deficits in patients with Huntington's disease or Alzheimer's disease: evidence from the CVLT-II. *J Clin Exp Neuropsychol.* (2008) 30:463–70. doi: 10.1080/13803390701531912
29. Strauss E, Sherman EMS, Spreen O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. 3rd ed. Oxford: Oxford University Press (2006).
30. Clark LR, Stricker NH, Libon DJ, Delano-Wood L, Salmon DP, Delis DC, et al. Yes/No versus forced-choice recognition memory in mild cognitive impairment and Alzheimer's disease: patterns of Impairment and associations with dementia severity. *J Clin Neuropsychol.* (2012) 26:1201–16. doi: 10.1080/13854046.2012.728626
31. Matura S, Prvulovic D, Butz M, Hartmann D, Sepanski B, Linnemann K, et al. Recognition memory is associated with altered resting-state functional connectivity in people at genetic risk for Alzheimer's disease. *Eur J Neurosci.* (2014) 40:3128–35. doi: 10.1111/ejn.12659
32. Murdock BB. The serial position effect of free recall. *J Exp Psychol.* (1962) 64:482–8. doi: 10.1037/h0045106
33. Bayley PJ, Salmon DP, Bond MW, Bui BK, Olichney J, Delis DC, et al. Comparison of the serial position effect in very mild Alzheimer's disease, mild Alzheimer's disease, and amnesia associated with electroconvulsive therapy. *J Int Neuropsychol Soc.* (2000) 6:290–8. doi: 10.1017/S1355617700633040
34. Howieson DB, Mattek N, Seyle AM, Dodge HH, Wasserman D, Zitzelberger T, et al. Serial position effects in mild cognitive impairment. *J Clin Exp Neuropsychol.* (2011) 33:292–9. doi: 10.1080/13803395.2010.516742
35. Jones SN, Greer AJ, Cox DE. Learning characteristics of the CERAD word list in an elderly VA sample. *Appl Neuropsychol.* (2011) 18:157–63. doi: 10.1080/09084282.2011.595441
36. Bruno D, Reichert C, Pomara N. The recency ratio as an index of cognitive performance and decline in elderly individuals. *J Clin Exp Neuropsychol.* (2016) 38:967–73. doi: 10.1080/13803395.2016.1179721
37. Bruno D, Kosciak RL, Woodard JL, Pomara N, Johnson SC. The recency ratio as predictor of early MCI. *Int Psychogeriatr.* (2018) 30:1883–8. doi: 10.1017/S1041610218000467
38. Tomadesso C, Gonneaud J, Egret S, Perrotin A, Pèlerin A, Flores R, et al. Is there a specific memory signature associated with Aβ-PET positivity in patients with amnesic mild cognitive impairment? *Neurobiol Aging.* (2019) 77:94–103. doi: 10.1016/j.neurobiolaging.2019.01.017
39. Bondi MW, Edmonds EC, Jak AJ, Clark LR, Delano-Wood L, McDonald CR, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimer's Dis.* (2014) 42:275–89. doi: 10.3233/JAD-140276
40. Jack CR Jr, Knopman DS, Weigand SD, Wiste HJ, Vemuri P, Lowe V, et al. An operational approach to national institute on aging-Alzheimer's association criteria for preclinical Alzheimer disease. *Ann Neurol.* (2012) 71:765–75. doi: 10.1002/ana.22628
41. Thomas KR, Edmonds EC, Eppig J, Salmon DP, Bondi MW, Alzheimer's Disease Neuroimaging Initiative. Using neuropsychological process scores to identify subtle cognitive decline and predict progression to mild cognitive impairment. *J Alzheimer's Dis.* (2018) 64:195–204. doi: 10.3233/JAD-180229
42. Knopman DS, Jones DT, Greicius MD. Failure to demonstrate efficacy of aducanumab: an analysis of the EMERGE and ENGAGE trials as reported by Biogen, December 2019. *Alzheimer's Dement.* (2021) 17:696–701. doi: 10.1002/alz.12213
43. Crane PK, Carle A, Gibbons LE, Insel P, Mackin RS, Gross A, et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav.* (2012) 6:502–16. doi: 10.1007/s11682-012-9186-z
44. Wang J, Logovinsky V, Hendrix SB, Stanworth SH, Perdomo C, Xu L, et al. ADCOMS: a composite clinical outcome for prodromal Alzheimer's disease trials. *J Neurol Neurosurg Psychiatry.* (2016) 87:993–9. doi: 10.1136/jnnp-2015-312383
45. Thomas KR, Eppig J, Edmonds EC, Jacobs DM, Libon DJ, Au R, et al. Word-list intrusion errors predict progression to mild cognitive impairment. *Neuropsychology.* (2018) 32:235–45. doi: 10.1037/neu0000413
46. Trenkle DL, Shankle WR, Azen SP. Detecting cognitive impairment in primary care: performance assessment of three screening instruments. *J Alzheimer's Dis.* (2007) 11:323–35. doi: 10.3233/JAD-2007-11309
47. Shankle WR, Mangrola T, Chan T, Hara J. Development and validation of the memory performance index: reducing measurement error in recall tests. *Alzheimer's Dement.* (2009) 5:295–306. doi: 10.1016/j.jalz.2008.11.001
48. Alexander GE, Satalich TA, Shankle WR, Batchelder WH. A cognitive psychometric model for the psychodiagnostic assessment of memory-related deficits. *Psychol Assess.* (2016) 28:279–93. doi: 10.1037/pas0000163
49. Lee MD, Bock JR, Cushman I, Shankle WR. An application of multinomial processing tree models and Bayesian methods to understanding memory impairment. *J Math Psychol.* (2020) 95:102328. doi: 10.1016/j.jmp.2020.102328
50. Lasserre JA, Bishop CM, Minka TP. Principled hybrids of generative and discriminative models. In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. New York, NY, USA. (2006).

**Conflict of Interest:** JB, JH, and DF are employees of Embic Corporation. JR is a consultant to Embic Corporation.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Bock, Russell, Hara and Fortier. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.