Preliminary Development of an Embedded Measure of Effort on the Memory Module of the Neuropsychological Assessment Battery (NAB)

Alexandra Rodriguez[1], Sierra Iwanicki, PhD[1], Alicia Carrillo, BS[1], David M. Lechuga, PhD[2], & Lisa Fasnacht-Hill, PhD[2]

[1]PAR, Inc.; [2]The Neurobehavioral Clinical and Counseling Center

PAR

Objective

Evaluation of performance validity is recognized as a necessary component of neuropsychological assessments recommended by both the American Academy of Clinical Neuropsychology (Heilbronner et al., 2009) and the National Academy of Neuropsychology (Bush et al., 2005). Without the use of performance validity tests (PVTs), accurate interpretation of test results obtained in a neuropsychological evaluation may be compromised. Traditionally, practitioners have relied on stand-alone measures of validity (e.g., Test of Memory Malingering [TOMM], Victoria Symptom Validity Test [VSVT], Validity Indicator Profile [VIP]), however there are several advantages to an embedded measure. They are efficient (i.e., they do not require additional time or testing) and can be applied retrospectively to previous testing (e.g., records review). Moreover, they allow for measurement of effort continuously throughout the battery (as effort may change during the course of the evaluation) and they are not recognizable as measures of effort (so they are less susceptible to coaching).

The Neuropsychological Assessment Battery (NAB) is an integrated neuropsychological battery used to assess cognitive skills in adults (Stern & White, 2003). A recent study developed two embedded PVTs within the NAB Screening Module using the Attention and Executive Functioning modules (Lace et al., 2021). Subtests that are most susceptible to poor effort (e.g., Digit Span and List Recognition) have been examined and used to develop an internal validity indicator within the Attention and Memory Modules of the NAB (Silverberg et al., 2007; Varela et al., 2021). However, further research focusing on the development of embedded PVTs within the main NAB modules (i.e., Memory, Attention, Language, Executive Functions, and Spatial) is needed (Varela et al., 2021). The purpose of this study was to utilize stand-alone PVTs to interpret suboptimal effort on the Memory Module of the NAB.

Method

Participants

This study utilized archival data from 407 adult civil litigants referred for a neuropsychological evaluation at a private practice clinic in the western United States. Participants with noncredible performance (n = 47) were defined as those who failed two or more PVTs. Participants with credible performance (n = 259) were defined as those who did not fail any PVTs. Due to ambiguity associated with classifying performance for failing one PVT, these individuals (n = 101) were excluded, resulting in a final sample of 306 individuals. The final sample was 56% male (n = 171) with ages ranging from 18 to 85 years with a mean age of 42.7 years (SD = 15.8). Ethnicity and handedness were not specified for 63.4% (n = 194) and 63.7% (n = 195) of participants. Education for the total sample was reported as 7.5% (n = 23) for less than high school (HS) diploma, 26.5% (n = 81) for HS diploma (or equivalent), 28.8% (n = 82) for some college, and 26.8% (n = 82) for those with a bachelor's degree or greater.

Measures

NAB Memory Module

- List Learning A
  - Immediate Recall
  - Short Delayed Recall
  - Long Delayed Recall
  - Discriminability Index
- Shape Learning
  - Immediate Recognition
- Delayed Living Memory
  - Immediate Recall
  - Delayed Recall
- Name/Address/Phone
  - Delayed Recall

Performance Validity Tests

- Dot Counting Test (DCT) (DCT; Boone et al., 2002b)
  - Invalid score cutoff: e-score ≥ 13.8
- b test (Boone et al., 2002a)
  - Invalid score cutoff: e-score ≥ 90
- Word Memory Test (WMT; Green, 2003)
  - Invalid cutoff: Immediate Recognition (IR), Delayed Recognition (DR) or Consistency Multiple Choice (MC) subtest score ≤ 82.5% or MC score of ≤ 70%, or Paired Associates (PA) subtest score of ≤ 60%

NAB Neuropsychological Assessment Battery®

Learn more

Procedure

Means and standard deviations were calculated for each subtest of the Memory Module for both credible and noncredible groups. Independent-samples t tests were conducted to examine mean differences on various subtests. Receiver-operating characteristic (ROC) curves were used to analyze the predictive accuracy of subtests of the NAB Memory Module. ROC curves were assessed by calculating the area under the curve (AUC); cutoff scores to maximize sensitivity and specificity were calculated.

Results

There were statistically significant differences in raw scores for the credible and noncredible groups on all 39 Memory Modules subtest scores. However, only 8 subtest scores demonstrated suitability for ROC analyses (AUC ≥ .80): List Learning A Immediate Recall, List Learning A Short Delayed Recall, List Learning A Long Delayed Recall, List Learning A Discriminability Index, Shape Learning Immediate Recognition, Daily Living Memory Immediate Recall, Daily Living Memory Delayed Recall, and Name/Address/Phone Delayed Recall. Means and standard deviations for the credible and noncredible groups as well as results from independent-samples t tests for each subtest are listed in Table 1.

Table 1. Results from Independent-Samples *T* Tests

Raw score	Credible M (SD)	Noncredible M (SD)	t (df)	Sig	Cohen's d
List Learning A Immediate Recall	22.6 (4.9)	15.7 (4.6)	–8.89 (300)	< .001	1.45
List Learning A Short Delayed Recall	7.3 (2.5)	3.9 (2.1)	–8.46 (300)	< .001	1.47
List Learning A Long Delayed	7.1 (2.7)	3.5 (2.6)	–8.25 (300)	< .001	1.36
List Learning A Discriminability Index	8.6 (2.8)	4.4 (3.5)	–7.50 (56.05)	< .001	1.33
Shape Learning Immediate Recognition	17.8 (3.8)	12.4 (4.7)	–8.57 (300)	< .001	1.26
Daily Living Memory Immediate Recall	42.4 (5.7)	32.8 (8.4)	–7.46 (52.64)	< .001	1.34
Daily Living Memory Delayed Recall	13.8 (3.1)	8.6 (4.6)	–7.34 (52.48)	< .001	1.33
Name/Address/Phone Delayed Recall	5.9 (2.1)	3.0 (2.3)	–8.43 (300)	< .001	1.32

Note: For subtests where Levene's test indicated unequal variances, the degrees of freedom were adjusted.

The ROC curves for using selected subtests to predict credible versus noncredible performance are shown in Figure 1. The cutoff scores, AUCs, 95% CIs, sensitivities, and specificities obtained for the selected NAB Memory Modules subtests are reported in Table 2.

Figure 1. Receiver-Operating Characteristic Curves (ROC) for Prediction of Credible Versus Noncredible Performance Based on NAB Subtest Scores

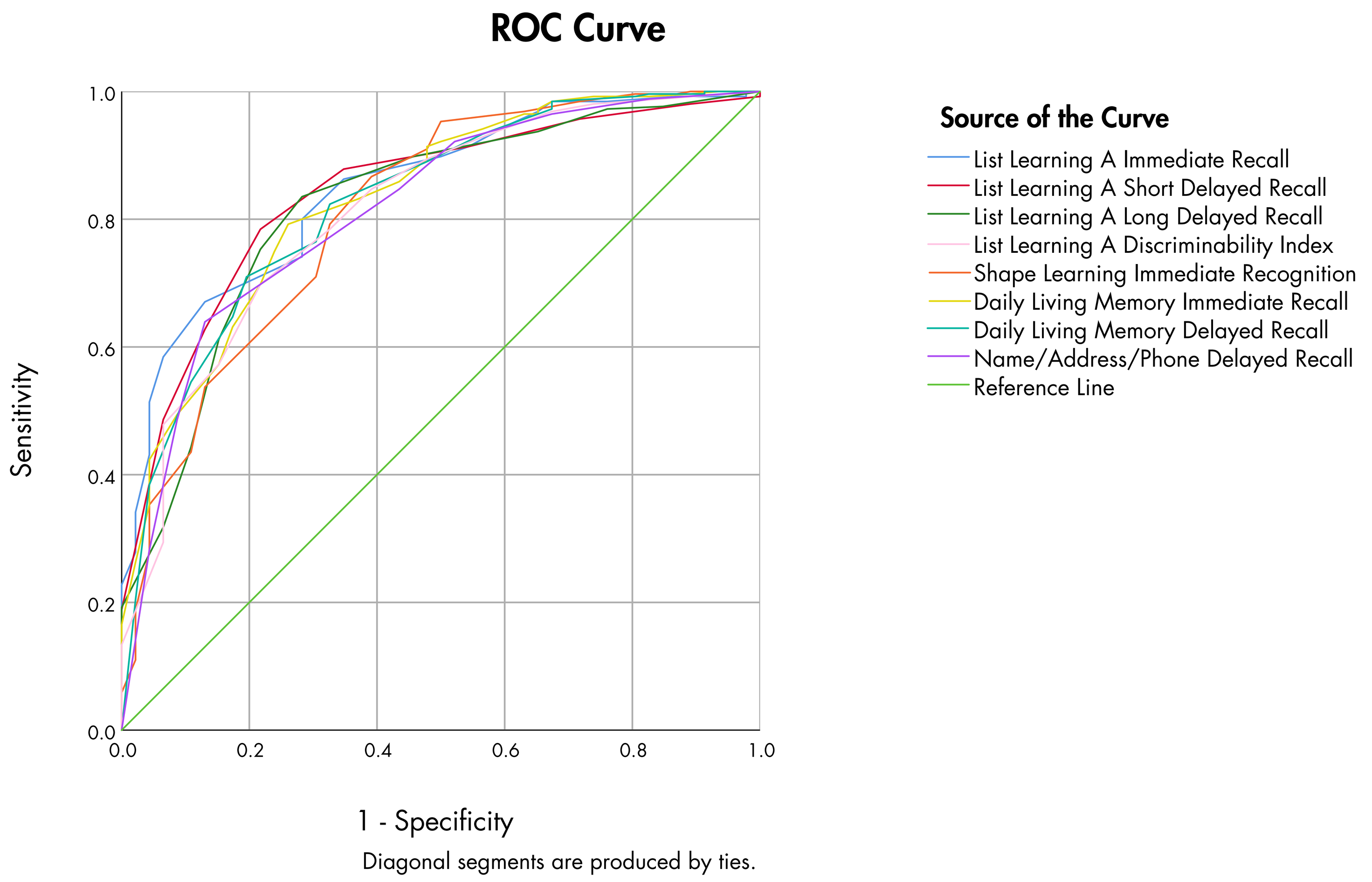

Table 2. Cutoff Scores, Area Under the Curve (AUC), Sensitivity, and Specificity for Each Subtest Score

Raw score	Cutoff score	AUC (CI 95%)	Sensitivity (%)	Specificity (%)
List Learning A Immediate Recall	≤ 18	.85 (.79–.90)	71.74	80.08
List Learning A Short Delayed Recall	≤ 5	.84 (.78–.90)	78.26	78.52
List Learning A Long Delayed Recall	≤ 4	.82 (.76–.89)	71.74	83.59
List Learning A Discriminability Index	≤ 7	.82 (.75–.88)	78.26	69.80
Shape Learning Immediate Recognition	≤ 15	.81 (.74–.88)	69.57	71.09
Daily Living Memory Immediate Recall	≤ 39	.83 (.77–.89)	76.09	74.61
Daily Living Memory Delayed Recall	≤ 11	.83 (.76–.89)	69.57	76.17
Name/Address/Phone Delayed Recall	≤ 4	.82 (.75–.88)	69.57	75.39

Conclusions

- These results provide preliminary psychometric evidence and clinical utility for use of various cutoff scores as a measure of embedded validity within the NAB Memory Module.
- Limitations of the current study include small sample size and the potential of artificially inflated classification statistics due to spectrum bias resulting from exclusion of participants who failed one PVT (Schroeder et al., 2019). Spectrum bias occurs when the spectrum of clinical manifestations in the data does not adequately reflect the spectrum in clinical practice (Park & Han, 2018).
- Further research should focus on pattern analysis employing a discriminant function analysis or logistic regression to create a scoring algorithm that includes multiple NAB subtests to better differentiate between known clinical groups (e.g., dementia and TBI patients) and suspected malingerers. Additionally, future research should focus on identifying suboptimal effort across multiple NAB modules.

References

Boone, K., Lu, P., & Herzberg, D. (2002a). *The b Test Manual.* Western Psychological Services.

Boone, K. B., Lu, P., & Herzberg, D. (2002b). *The Dot Counting Test.* Western Psychological Services.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology, 20*(4), 419–426. https://doi.org/fspx9h

Green, P. (2003). *Word Memory Test for Windows: User's Manual and Program.* Author.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G., Millis, S. R., & Participants, C. (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*(7), 1093–1129. https://doi.org/fpz64r

Lace, J. W., Grant, A. F., Ruppert, P., Kaufman, D. A. S., Teague, C. L., Lowell, K., & Gfeller, J. D. (2021). Detecting noncredible performance with the neuropsychological assessment battery, screening module: A simulation study. *The Clinical Neuropsychologist, 35*(3), 572–596. https://doi.org/gzsk

Park, S. H., & Han, K. (2018). Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology, 286*(3), 800–809. https://doi.org/gc46wr

Schroeder, R. W., Martin, P. K., Heinrichs, R. J., & Baade, L. E. (2019). Research methods in performance validity testing studies: Criterion grouping approach impacts study outcomes. *The Clinical Neuropsychologist, 33*(3), 466–477. https://doi.org/dq24

Silverberg, N. D., Wertheimer, J. C., & Fichtenberg, N. L. (2007). An effort index for the repeatable battery for the assessment of neuropsychological status (RBANS). *The Clinical Neuropsychologist, 21*(5), 841–854. https://doi.org/bfhkpn

Stern, R. A., & White, T. (2003). *Neuropsychological Assessment Battery: Administration, Scoring, and Interpretation manual.* PAR.

Varela, J. L., Ord, A. S., Phillips, J. I., Shura, R. D., & Sautter, S. W. (2021). The development and validation of the embedded validity indicator for the neuropsychological assessment battery. *Archives of Clinical Neuropsychology,* 1–13. https://doi.org/gzsm