Psychological Injury and Law (2023) 16:27–35
https://doi.org/10.1007/s12207-022-09443-3

Cross-Validation of Multiple Embedded Performance Validity Indices in the Rey Auditory Verbal Learning Test and Brief Visuospatial Memory Test-Revised in an Adult Attention Deficit/Hyperactivity Disorder Clinical Sample

Matthew S. Phillips[1] · Amanda M. Wisinger[1] · Franchezka T. Lapitan-Moore[1] · Jenna E. Ausloos-Lozano[1] · Hanaan Bing-Canar[1] · Nicole M. Durkin[1] · Gabriel P. Ovsiew[1] · Zachary J. Resch[1] · Kyle J. Jennette[1] · Jason R. Soble[1,2]

Received: 30 August 2021 / Accepted: 9 January 2022 / Published online: 18 January 2022

Abstract

This study cross-validated the Rey Auditory Verbal Learning Test (RAVLT) Effort Score (ES), RAVLT Forced Choice (FC), and Brief Visuospatial Memory Test-Revised (BVMT-R) Recognition Discrimination (RD) among a large, consecutive case series of 317 adults referred for outpatient neuropsychological evaluation of known or suspected Attention-Deficit/Hyperactivity Disorder (ADHD). Six independent criterion performance validity tests (PVTs) were used to establish valid ($n=280$) and invalid ($n=37$) performance groups. Among the valid group, 229 met *Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition* criteria for ADHD. Non-significant to small correlations emerged between ES, FC, RD, and the criterion PVTs, suggesting these measures are largely independent. The valid group performed significantly better than the invalid group across all three embedded PVTs, with medium to large effects ($d=0.66$–0.89). Receiver operator characteristic (ROC) curve analyses similarly revealed that all three PVTs accurately differentiated valid from invalid performance (areas under the curve: .63–.74) with the following optimal cut-scores: ES $\leq$ 12 (43% sensitivity/90% specificity), FC $\leq$ 14 (49% sensitivity/92% specificity), RD $\leq$ 5 (35% sensitivity/90% specificity). Results were identical between the overall sample of adult ADHD referrals and the subsample who met diagnostic criteria for ADHD. In sum, these memory-based RAVLT and BVMT-R PVTs were able to accurately identify invalid neuropsychological test performance among adults undergoing evaluation for ADHD, regardless of whether diagnostic criteria for ADHD was met. Although optimal cut-scores generally were consistent with prior literature, sensitivity was generally less robust than prior cross-validation studies with mixed clinical and medicolegal samples.

Keywords Performance Validity · ADHD · Assessment · Neuropsychology

Introduction

Regardless of referral reason, PVTs are a well-established component of neuropsychological assessment to ensure that test data are an accurate representation of a patient's true neurocognitive functioning. Unsurprisingly, current practice guidelines support the administration of multiple freestanding and embedded PVTs throughout the evaluation, which are ultimately aggregated together to provide a comprehensive evaluation of the validity of a patient's test performance (Boone, 2009, 2013; Bush et al., 2005; Larrabee, 2008; Sweet et al., 2021). Although practice standards call for routine validity assessment in all evaluations, it is particularly salient in populations found to have higher base rates of performance invalidity, such as adults presenting for evaluation of Attention Deficit/Hyperactivity Disorder (ADHD).

Reported base rates of performance invalidity among adult ADHD evaluations vary in the current literature, ranging from 13 to 50% (Suhr et al., 2008; Sullivan et al., 2007; White et al., 2020b), which is at least partially attributable

✉ Matthew S. Phillips
mxp9351@ego.thechicagoschool.edu

[1] Department of Psychiatry, University of Illinois College of Medicine, Chicago, IL, USA

[2] Department of Neurology, University of Illinois College of Medicine, Chicago, IL, USA

to differences in study designs and operational definitions of performance invalidity across studies. Additionally, Martin and Schroeder (2020) recently surveyed practicing neuropsychologists and documented a 20% reported invalidity base rate for adult ADHD evaluations. These varying base rates of observed and reported performance invalidity in the context of adult ADHD evaluations further complicates interpretation of positive and negative predictive values given the true base rate of performance invalidity. Nonetheless, considering the multiple potential external incentives associated with an ADHD diagnosis, including prescription of psychostimulant medication or provision of academic/testing accommodations (Advokat et al., 2008; Musso & Gouvier, 2014; Teter et al., 2005), the higher-than-average base rate of invalidity in this population is somewhat unsurprising. The dissociation of performance and symptom validity in adult ADHD populations (e.g., Leib et al., 2021; White et al., 2020b) further complicates accurate identification of invalid cognitive test performance and places greater onus on PVT findings. A final relevant consideration for performance validity assessment in ADHD relates to the fact that some of the most commonly administered PVTs assess (or at least appear to assess) memory (Martin et al., 2015). Given that memory dysfunction may not intuitively be considered a characteristic feature of ADHD, particularly by examinees, it is essential that such memory-based PVTs be cross-validated among those undergoing evaluation for ADHD to ensure that their sensitivity remains robust in this select clinical population. Along with robust sensitivity, showing adequate specificity (i.e., ≥ . 90; Babikian et al., 2006; Larrabee, 2008, 2012; Boone, 2011; Binder et al., 2014) of these PVTs also is critical to prevent those with valid performance being misclassified as invalid.

Given that memory is a cognitive domain that has been objectively shown to be impacted by ADHD in both individual studies (e.g., Keezer et al., 2021) and larger systematic reviews (Pievsky & McGrath, 2018), accurate assessment of learning/memory often is a critical component of a comprehensive neuropsychological evaluation in the context of ADHD. The (RAVLT; Rey, 1941; Schmidt, 1996) and Brief Visuospatial Memory Test-Revised (BVMT-R; Benedict, 1997) are tests of verbal and visual learning/memory, respectively, that are among the most commonly administered memory measures among practicing neuropsychologists (Rabin et al., 2016). In addition to being frequently used and well-researched memory tests, both measures contain embedded PVTs that have been cross-validated and found to accurately discriminate between valid and invalid test performances across a range of medicolegal and clinical populations (e.g., Boone et al., 2005; Bailey et al., 2018; Pliskin et al., 2021; Olsen et al., 2019; Poreh et al., 2016; Resch et al., 2020; Soble et al., 2021). Nonetheless, the accuracy of these previously developed RAVLT and BVMT-R PVTs has not been adequately cross-validated among adult patients referred for ADHD evaluations. If found to have similarly favorable sensitivity for detecting performance invalidity, these embedded RAVLT and BVMT-R PVTs may be particularly desirable to integrate into adult ADHD neuropsychological evaluations, given they do not necessitate the administration of a separate assessment measure and can thereby reduce total evaluation time, patient burden, and costs. Accordingly, this study aimed to evaluate the accuracy of the RAVLT Effort Score (ES), RAVLT Forced Choice (FC), and BVMT-R Recognition Discrimination (RD) embedded PVTs among a large sample of adults referred for outpatient neuropsychological evaluation of known or suspected ADHD.

Method

Participants

This cross-sectional study utilized data from a consecutive case series of 323 adult referrals for comprehensive outpatient neuropsychological evaluation. All patients were referred by their treating providers, most often primary care physicians or psychiatrists, specifically for purposes of differential diagnosis and treatment planning related to known or suspected ADHD. Six patients were missing a measure of interest for this study and were excluded, resulting in a final sample of 317. The overall sample was 61% female ($n = 194$) and 39% male ($n = 123$) with a mean age of 27.70 ($SD = 6.67$) and mean educational attainment of 15.71 years ($SD = 2.05$). Racial/ethnic composition was 46% White ($n = 147$), 24% Hispanic ($n = 75$), 15% Black ($n = 46$), 10% Asian ($n = 32$), and 5% other ($n = 17$). A majority of the sample (69%; $n = 220$) were actively enrolled college students at the time of their evaluation. Validity groups were established based on performance on six independent criterion PVTs (listed in Table 1) and current practice standards/empirical findings (e.g., Critchfield et al., 2019; Jennette et al., 2021; Larrabee, 2008; Rhoads et al., 2021; Sherman et al., 2020; Soble et al., 2020; Webber et al., 2020), such that those with one or fewer criterion PVT fails were classified as having valid test performance ($n = 280$), whereas those with two or more criterion PVT failures demonstrated invalid test performance ($n = 37$).

Among the 280 patients with valid test performance, 229 met formal *Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition* (DSM-5; American Psychiatric Association, 2013) diagnostic criteria for ADHD, whereas the remaining 51, despite having attention complaints, did not meet criteria for ADHD. All the ADHD diagnoses were rendered by a board-certified clinical neuropsychologist and were based on a clinic standard diagnostic protocol that included a medical/psychiatric record review, including previous ADHD

Table 1 Independent criterion performance validity tests used to establish validity groups

Performance validity test	Failure cut-score(s)	Sensitivity/specificity	Reference	Sample failure rate
Dot Counting Test	E-score ≥ 14	88%/96%	Boone et al. (2002)	39/317 (12%)
Rey 15-Item Test	Recall + Recognition ≤ 22	61%/90%	Poynter et al. (2019)	9/317 (3%)
Stroop Color Word Test	Word Reading T-score ≤ 21	54%/91%	White et al. (2020a, b)	23/317 (7%)
Trail Making Test	Part A T-score ≤ 34	35%–53%/89%–95%	White et al. (2020a, b) Ashendorf et al. (2017)	36/317 (11%)
Verbal Fluency	F/A/S T-score ≤ 32	42%/90%	White et al. (2020a, b)	26/317 (8%)
WAIS-IV Digit Span	Reliable Digit Span ≤ 7	22%–58%/85%–94%	Marshall et al. (2010) Schroeder et al. (2012)	33/317 (10%)

N = 317. *WAIS-IV* Wechsler Adult Intelligence Scale-Fourth Edition

testing/evaluations, when available; patient completion of a detailed history and symptom questionnaire; a semi-structured clinical interview covering DSM-5 criteria as well as relevant medical, psychiatric, developmental, and psychosocial history/symptoms; an ADHD symptom report measure with embedded symptom validity scales (i.e., Clinical Assessment of Attention Deficit-Adult; Bracken & Boatwright, 2005); and a uniform neuropsychological test battery which included the RAVLT, BVMT-R, and the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008).

Measures

Brief Visuospatial Memory Test-Revised

The BVMT-R (Benedict, 1997) is a test of visuospatial learning and memory in which the patients are shown an array of six figures that they must reproduce over three learning trials, followed by delayed recall and recognition trials. Bailey et al. (2018) initially validated the Recognition Discrimination (RD; Hits — False Positives) as an embedded PVT, which has subsequently been cross-validated across diverse clinical populations (e.g., Olsen et al., 2019; Pliskin et al., 2021; Resch et al., 2020).

Rey Auditory Verbal Learning Test

The RAVLT (Rey, 1941; Schmidt, 1996) is a test of verbal/auditory learning and memory in which patients are presented a list of 15 words over five learning trials followed by short- and long-delay free recall trials and a recognition paradigm. Boone et al. (2005) initially validated an Effort Score (E-score) based on recognition performance (Hits — False Positives + the number of the first five list words correctly recognized), which has also received cross-validation support (e.g., Pliskin et al., 2021; Whitney & Davis, 2015). More recently, Poreh et al. (2016) developed a forced-choice addendum to the RAVLT, which has also shown promise as a robust PVT on cross-validation (Soble et al., 2021). For this study, both the RAVLT ES and FC scores were examined.

Data Analyses

Spearman correlations examined relationships among the BVMT-R RD, RAVLT FC, RAVLT ES, as well as their associations with the criterion PVTs (Table 2). Descriptive statistics were calculated for each test PVT (i.e., RAVLT ES, RAVLT FC, BVMT-R RD), and non-parametric Mann–Whitney *U*-tests assessed for significant performance differences between the valid and invalid groups. Finally, receiver operating characteristics (ROC) curve analyses then investigated each PVT's overall classification accuracy and, for PVTs with a significant area under the curve (AUC), determined optimal cut-scores that maximize sensitivity and specificity. Positive (PPV) and negative (NPV) predictive values were also calculated at multiple invalidity base rates. Classification accuracy was interpreted as low (0.50–0.69), acceptable (0.70–0.79), excellent (0.80–0.89), or outstanding (≥ 0.90) based on AUCs (Hosmer et al., 2013). All the analyses were first conducted for the overall sample (i.e., ADHD referrals) and then repeated with the subsample of patients in the valid group who met DSM-5 diagnostic criteria for ADHD to examine if the three embedded PVTs functioned similarly in a more homogenous ADHD sample. Finally, ROC analyses were rerun for the overall sample excluding those with one criterion PVT failure to examine for potential differences using a more stringent criterion grouping method (i.e., 0 vs. 2 + PVT fails).

Results

The RAVLT ES, RAVLT FC, and BVMT-R RD had non-significant to small correlations with the criterion PVTs as well as with each other, indicating that these measures were largely independent. Performance on the RAVLT and

Table 2 Intercorrelations between performance validity tests among the valid group

	DCT	FAS	RDS	RFIT	Stroop	TMT-A	RAVLT ES	RAVLT FC	BVMT RD
DCT	-	−.09	−.17**	−.17**	−.31**	−.18**	−.22**	−.09	−.08
FAS	-	-	.18**	−.05	.21**	.17**	−.01	.06	.18**
RDS	-	-	-	.16**	.18**	−.03	.08	.02	.05
RFIT	-	-	-	-	−.01	.02	.20**	.11	.09
Stroop	-	-	-	-	-	.23**	.04	−.02	.01
TMT-A	-	-	-	-	-	-	.09	.12*	.12*
RAVLT ES	-	-	-	-	-	-	-	.34**	.09
RAVLT FC	-	-	-	-	-	-	-	-	.23**
BVMT RD	-	-	-	-	-	-	-	-	-

$n = 280$. *DCT* Dot Counting Test, *FAS* Verbal Fluency Test, *RDS* Reliable Digit Span, *RFIT* Rey-15 Item Test/Recognition, *TMT-A* Trail Making Test-Part A, *RAVLT-ES* Rey Auditory Verbal Learning Test Effort Score, *FC* RAVLT Forced Choice, *BVMT-RD* Brief Visuospatial Memory Test-Recognition Discrimination

$^{*}p < .05$, $^{**}p < .01$

BVMT-R embedded validity indicators by validity group is presented in Table 3. Comparison between the valid and invalid groups within the overall sample found that the valid group performed significantly better (i.e., higher scores) across all three indices relative to the invalid group, with large effects. When examining the performance of the valid and invalid groups among the ADHD subsample, these findings also held constant and were comparable to the results with the entire sample.

Among the entire sample (Table 4), all three embedded PVTs significantly differentiated valid from invalid performance. Classification accuracies ranged from low for the BVMT-R RD (AUC = 0.63) to acceptable for the RAVLT ES (AUC = 0.70) and FC (AUC = 0.74). The BVMT-R RD had 35% sensitivity/90% specificity at an optimal cut-score of ≤ 5. The RAVLT FC had 49% sensitivity/92% specificity at an optimal cut-score of ≤ 14, and the RAVLT ES demonstrated 43% sensitivity/90% specificity at an optimal cut-score of ≤ 12. See Table 4 for PPV/NPV values at 10%, 20%, and 30% base rates of performance invalidity.

When comparing the valid groups to the invalid groups among the ADHD subsample (Table 5), results were largely identical to those of the overall sample. Notably, classification accuracies for all three embedded PVTs were statistically significant with the BVMT-R RD having lower classification accuracy (AUC = 0.63) relative to the RAVLT FC (AUC = 0.70) and ES (AUC = 0.74). Optimal cut-scores remained consistent with those of the overall sample (BVMT-R RD ≤ 5; RAVLT FC ≤ 14; RAVLT ES ≤ 12) and had similar sensitivity and specificity values. See Table 5 for PPV/NPV values at 10%, 20%, and 30% invalidity base rates for this subsample.

Finally, as shown in Table 6, results were nearly identical when ROC analyses were rerun excluding those with one criterion PVT failure. As such, using 0–1 criterion PVT failure did not yield meaningful different classification accuracy or associated psychometrics than using a more stringent 0 vs. ≥ 2 criterion classification method to establish validity groups.

Table 3 Performance on Rey Auditory Verbal Learning Test and Brief Visuospatial Memory Test-Revised embedded performance indictors by validity group

Overall sample/invalid	**Valid (n = sample/invalid 280) M (SD)**	**Invalid (n = sample/invalid 37) M (SD)**	U	d
RAVLT ES	17.17 (3.29)	13.70 (4.39)	2678.00***	0.89
RAVLT FC	14.87 (0.52)	14.19 (1.00)	3066.00***	0.85
BVMT-R RD	5.88 (0.40)	5.30 (1.18)	3813.00***	0.66
ADHD subsample/invalid	**Valid (n = 229) M (SD)**	**Invalid (n = 37) M (SD)**		
RAVLT ES	17.09 (3.24)	13.70 (4.39)	2242.50***	0.88
RAVLT FC	14.86 (0.49)	14.19 (1.00)	2532.50***	0.85
BVMT-R RD	5.87 (0.43)	5.30 (1.18)	3126.00***	0.64

RAVLT-ES Rey Auditory Verbal Learning Test Effort Score, *FC* RAVLT Forced Choice, *BVMT-RD* Brief Visuospatial Memory Test-Recognition Discrimination

$^{***}p < .001$

Table 4 Accuracy of the Rey Auditory Verbal Learning Test and Brief Visuospatial Memory Test-Revised embedded performance indicators for detecting invalidity for the overall sample

	Cutoff	SN (%)	SP (%)	30% Base Rate PPV	30% Base Rate NPV	20% Base Rate PPV	20% Base Rate NPV	10% Base Rate PPV	10% Base Rate NPV
RAVLT ES									
AUC = .74* **	≤9	.19	.96	0.76	0.64	0.67	0.73	0.54	0.83
	≤10	.27	.94	0.75	0.66	0.66	0.75	0.53	0.84
	≤11	.35	.92	0.76	0.68	0.67	0.77	0.54	0.85
	≤12	**.43**	**.90**	0.74	0.70	0.65	0.79	0.52	0.86
	≤13	.46	.88	0.72	0.71	0.62	0.79	0.49	0.87
	≤14	.49	.84	0.67	0.71	0.57	0.79	0.43	0.87
RAVLT FC									
AUC = .70* **	≤12	.08	.99	0.84	0.62	0.77	0.72	0.67	0.81
	≤13	.24	.97	0.84	0.66	0.77	0.75	0.67	0.84
	≤14	**.49**	**.92**	0.80	0.73	0.72	0.81	0.60	0.88
BVMT-R RD									
AUC = .63 **	≤2	.05	.99	0.77	0.61	0.68	0.71	0.56	0.81
	≤3	.11	.99	0.88	0.63	0.83	0.72	0.73	0.82
	≤4	.19	.99	0.93	0.65	0.89	0.74	0.83	0.83
	≤5	**.35**	**.90**	0.70	0.68	0.60	0.76	0.47	0.85

$N = 317$. *RAVLT-ES* Rey Auditory Verbal Learning Test Effort Score, *FC* RAVLT Forced Choice, *BVMT-RD* Brief Visuospatial Memory Test-Recognition Discrimination, *AUC* area under the curve, *SN* sensitivity, *SP* specificity, *PPV* positive predictive value, *NPV* negative predictive value. Bolded scores denote optimal cut-scores for detecting invalid performance

$^{**}p < .01$, $^{***}p < .001$

Table 5 Accuracy of the Rey Auditory Verbal Learning Test and Brief Visuospatial Memory Test-Revised embedded performance indicators for detecting invalidity for the subsample of the valid group diagnosed with Attention Deficit/Hyperactivity Disorder

	Cutoff	SN (%)	SP (%)	30% Base rate PPV	30% Base rate NPV	20% Base rate PPV	20% Base rate NPV	10% Base rate PPV	10% Base rate NPV
RAVLT ES									
AUC = .74* **	≤9	.19	.96	0.67	0.73	0.54	0.83	0.35	0.91
	≤10	.27	.94	0.66	0.75	0.53	0.84	0.33	0.92
	≤11	.35	.92	0.65	0.77	0.52	0.85	0.33	0.93
	≤12	**.43**	**.90**	0.65	0.79	0.52	0.86	0.32	0.93
	≤13	.46	.88	0.62	0.79	0.49	0.87	0.30	0.94
	≤14	.49	.85	0.58	0.80	0.45	0.87	0.27	0.94
RAVLT FC									
AUC = .70* **	≤12	.08	.99	0.84	0.62	0.77	0.72	0.67	0.81
	≤13	.24	.97	0.84	0.66	0.77	0.75	0.67	0.84
	≤14	**.49**	**.91**	0.78	0.73	0.70	0.81	0.58	0.88
BVMT-R RD									
AUC = .63*	≤2	.05	.99	0.68	0.71	0.56	0.81	0.36	0.90
	≤3	.11	.99	0.83	0.72	0.73	0.82	0.55	0.91
	≤4	.19	.99	0.89	0.74	0.83	0.83	0.68	0.92
	≤5	**.35**	**.90**	0.60	0.76	0.47	0.85	0.28	0.93

$N = 266$. *RAVLT-ES* Rey Auditory Verbal Learning Test Effort Score, *FC* RAVLT Forced Choice, *BVMT-RD* Brief Visuospatial Memory Test-Recognition Discrimination, *AUC* area under the curve, *SN* sensitivity, *SP* specificity, *PPV*, positive predictive value, *NPV* negative predictive value. Bolded scores denote optimal cut-scores for detecting invalid performance

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 6 Accuracy of the Rey Auditory Verbal Learning Test and Brief Visuospatial Memory Test-Revised embedded performance indicators for detecting invalidity for the overall sample excluding patients with one criterion performance validity test failure

	Cutoff	SN (%)	SP (%)	30% Base rate		20% Base rate		10% Base rate	
				PPV	NPV	PPV	NPV	PPV	NPV
RAVLT ES									
AUC = .76* **	≤9	.19	.96	0.67	0.73	0.54	0.83	0.35	0.91
	≤10	.27	.95	0.70	0.75	0.57	0.84	0.38	0.92
	≤11	.35	.93	0.68	0.77	0.56	0.85	0.36	0.93
	≤12	**.43**	**.92**	0.70	0.79	0.57	0.87	0.37	0.94
	≤13	.46	.90	0.66	0.80	0.53	0.87	0.34	0.94
	≤14	.49	.86	0.60	0.80	0.47	0.87	0.28	0.94
RAVLT FC									
AUC = .71* **	≤12	.08	.99	0.77	0.72	0.67	0.81	0.47	0.91
	≤13	.24	.98	0.84	0.75	0.75	0.84	0.57	0.92
	≤14	**.49**	**.93**	0.75	0.81	0.64	0.88	0.44	0.94
BVMT-R RD									
AUC = .64*	≤2	.05	.99	0.68	0.71	0.56	0.81	0.36	0.90
	≤3	.11	.99	0.83	0.72	0.73	0.82	0.55	0.91
	≤4	.19	.99	0.89	0.74	0.83	0.83	0.68	0.92
	≤5	**.35**	**.91**	0.63	0.77	0.49	0.85	0.30	0.93

N = 211. *RAVLT-ES* Rey Auditory Verbal Learning Test Effort Score, *FC* RAVLT Forced Choice, *BVMT-RD* Brief Visuospatial Memory Test-Recognition Discrimination, *AUC* area under the curve, *SN* sensitivity, *SP* specificity, *PPV* positive predictive value, *NPV* negative predictive value

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Discussion

This study cross-validated three previously identified, embedded memory-based PVTs derived from the BVMT-R and RAVLT in a sample of adults referred for neuropsychological evaluations of known or suspected ADHD. Among the total sample (i.e., ADHD referrals), results revealed that all three embedded PVTs (i.e., RAVLT FC, RAVLT ES, BVMT-R RD) significantly differentiated valid from invalid performance. Comparatively, the RAVLT FC (AUC = 0.70) and ES (AUC = 0.74) had acceptable classification accuracies at optimal cut-scores of ≤ 14 (49% sensitivity/92% specificity) and ≤ 12 (43% sensitivity/90% specificity), respectively. In contrast, the BVMT-R RD had lower classification accuracy (AUC = 0.63) at an optimal cut-score of ≤ 5 (35% sensitivity/90% specificity). Among the subsample of patients who met DSM-5 diagnostic criteria for ADHD, the results were nearly identical to those of the overall sample, with all three embedded indices significantly distinguishing valid from invalid performance and producing similar classification accuracies, optimal cut-scores, sensitivities, and specificities. Therefore, results indicate that these cut-scores can be used in the context of ADHD evaluation, regardless of the presence/absence of ADHD. Finally, results were remarkably stable irrespective of whether patients in the valid group had zero or one criterion PVT fail, which is consistent with previous literature in other clinical samples (Abramson et al., 2021; Jennette et al., 2021; Rhoads et al., 2021).

Findings from this study provide further support to the growing body of literature highlighting the utility of the BVMT-R RD, RAVLT FC, and RAVLT ES as embedded PVTs (e.g., Bailey et al., 2018; Pliskin et al., 2021; Resch et al., 2020; Soble et al., 2021), such that these indices can accurately distinguish valid from invalid performance among ADHD referrals at roughly the same cut-scores identified in prior studies. However, compared to other clinical samples, all three embedded indices demonstrated relatively lower sensitivities in this sample of ADHD referrals. In prior studies, BVMT-R RD sensitivities generally ranged from 40 to 50% across samples (Bailey et al., 2018; Resch et al., 2020). Similarly, sensitivities of 66–82% have been found for the RAVLT FC (Soble et al., 2021) and 74–80% for RAVLT ES (Boone et al., 2005; Pliskin et al., 2021). Therefore, despite lower sensitivities across measures, the overall pattern of findings with the BVMT-R RD having lower classification accuracy and sensitivity than the RAVLT ES and FC indices matches the extant literature. Ultimately, findings suggest that, although these embedded indices certainly have clinical utility in detecting invalid performance, they may not be as sensitive among patients referred for ADHD evaluation. In particular, as the true base rate of performance invalidity in ADHD remains somewhat unclear, incorporating PPV/NPV values at various base rates is critical, given these values can vary considerably as a function of the base rate of invalidity, as was also the case with this study.

The current study demonstrated several noteworthy methodological strengths. First, the study comprised a large, demographically diverse sample. Second, multiple memory- and non-memory-based criterion PVTs were employed to independently establish validity classifications for the sample. Third, ADHD diagnoses were obtained using a standardized, multimodal diagnostic assessment protocol to facilitate accurate diagnosis. Finally, study analyses were replicated with the subsample of patients in the valid group who met DSM-5 diagnostic criteria for ADHD to establish that the three embedded PVTs of interest functioned comparably regardless of whether patients ultimately met the diagnostic criteria for ADHD. Despite these strengths, study findings must be interpreted while considering relevant limitations. Although typical of adult ADHD assessments, corroborating reports of ADHD symptomatology from informants were largely not available. Furthermore, the mean educational attainment (i.e., 15.71 years) from the sample was high and, as a result, study findings may not generalize to individuals with lower educational attainment. Future research would benefit from replication in a sample with greater educational diversity. Additionally, several of the criterion PVTs utilized to establish participants' validity status were embedded measures. Although embedded PVTs can be less psychometrically robust relative to freestanding PVTs (e.g., Bain et al., 2021; Ovsiew et al., 2020; Pliskin et al., 2021; Webber & Soble, 2018), the six independent criterion PVTs that were used in the neuropsychological battery have been found to have adequate psychometric properties via independent cross-validation studies. Given it has been shown that aggregating validity indices enhance their ability to detect invalidity (e.g., Larrabee, 2008), elevated false positive or false negative classification errors in this study are less likely. Nonetheless, as no criterion approach for determining validity status is perfect, the possibility of some patients with invalid performance being inaccurately classified as valid (i.e., false negatives) exists. As such, future research in this population would benefit from incorporating more freestanding criterion PVTs.

Taken together, current findings support the utility of the BVMT-R RD, RAVLT FC, and RAVLT ES embedded indices for distinguishing valid from invalid performance among ADHD referrals. Scores of ≤ 5 on the BVMT-R RD, ≤ 14 on the RAVLT FC, and ≤ 12 on the RAVLT ES were identified as optimal cut scores for both the overall sample of ADHD referred patients and the subset with positive ADHD diagnoses. Consistent with prior literature, the BVMT-R RD evidenced lower classification accuracy and sensitivity than the RAVLT FC and ES. However, all three indices evidenced lower sensitivities to those reported in other types of clinical samples. This discrepancy may be due, at least in part, to the memory-based nature of the BVMT-R and RAVLT as individuals referred for ADHD evaluation may be more likely to feign attention difficulties since ADHD is not traditionally thought of as a memory-based disorder.

References

Abramson, D. A., White, D. J., Rhoads, T., Carter, D. A., Hansen, N. D., Resch, Z. J., Jennette, K. J., Ovsiew, G. P., & Soble, J. R. (2021). Cross-validating the Dot Counting Test among an adult ADHD clinical sample and analyzing the effect of ADHD subtype and comorbid psychopathology. *Assessment*. https://doi.org/10.1177/10731911211050895

Advokat, C. D., Guidry, D., & Martino, L. (2008). Licit and illicit use of medications for attention-deficit hyperactivity disorder in undergraduate college students. *Journal of American College Health, 56*(6), 601–606. https://doi.org/10.3200/JACH.56.6.601-606

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., DSM-5). American Psychiatric Publishing.

Ashendorf, L., Clark, E. L., & Sugarman, M. A. (2017). Performance validity and processing speed in a VA polytrauma sample. *The Clinical Neuropsychologist, 31*(5), 857–866. https://doi.org/10.1080/13854046.2017.1285961

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist, 20*(1), 145–159. https://doi.org/10.1080/13854040590947362

Bailey, K. C., Soble, J. R., Bain, K. M., & Fullen, C. (2018). Embedded performance validity tests in the Hopkins verbal learning test-Revised and the brief visuospatial memory test-revised: A replication study. *Archives of Clinical Neuropsychology, 33*(7), 895–900. https://doi.org/10.1093/arclin/acx111

Bain, K. M., Soble, J. R., Webber, T. A., Messerly, J. M., Bailey, K. C., Kirton, J. W., & McCoy, J. J. M. (2021). Cross-validation of three advanced clinical solutions performance validity tests: Examining combinations of measures to maximize classification of invalid performance. *Applied Neuropsychology: Adult, 28*(1), 24–34. https://doi.org/10.1080/23279095.2019.1585352

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF: Manual for administration, scoring, and interpretation*. University of Minnesota Press.

Benedict, R. H. B. (1997). *Brief Visuospatial Memory Test—Revised*. Odessa, FL: Psychological Assessment Resources.

Binder, L. M., Larrabee, G. J., & Millis, S. R. (2014). Intent to fail: Significant testing of forced choice test results. *The Clinical Neuropsychologist, 28*(8), 1366–1375. https://doi.org/10.1080/13854046.2014.978383

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist, 23*(4), 729–741. https://doi.org/10.1080/13854040802427803

Boone, K. B. (2011). Clarification or confusion? A review of Rogers, Bender, and Johnson's a critical analysis of the MND criteria for feigned cognitive impairment: Implications for forensic practice and research. *Psychological Injury and Law, 4*(2), 157–162. https://doi.org/10.1007/s12207-011-9106-3

Boone, K. B. (2013). *Clinical practice of forensic neuropsychology: An evidenced-based approach.* Guilford Press.

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

Boone, K. B., Lu, P., & Wen, J. (2005). Comparison of various RAVLT scores in the detection of noncredible memory performance. *Archives of Clinical Neuropsychology, 20*(3), 301–319. https://doi.org/10.1016/j.acn.2004.08.001

Bracken, B. A., & Boatwright, B. S. (2005). *Clinical assessment of attention deficit – child and adult: Examiner's manual*. Psychological Assessment Resources.

Bush, S. S., Ruff, R. M., Troster, 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/10.1016/j.acn.2005.02.002

Critchfield, E., Soble, J. R., Marceaux, J. C., Bain, K. M., Bailey, K. C., Webber, T. A., Alverson, W. A., Messerly, J., González, D. A., & O'Rourke, J. J. F. (2019). Cognitive impairment does not cause performance validity failure: Analyzing performance patterns among unimpaired, impaired, and noncredible participants across six tests. *The Clinical Neuropsychologist, 33*(6), 1083–1101. https://doi.org/10.1080/13854046.2018.1508615

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression (3rd Ed.)*. John Wiley & Sons.

Jennette, K. J., Williams, C. P., Resch, Z. J., Ovsiew, G. P., Durkin, N. M., O'Rourke, J. J. F., Marceaux, J. C., Critchfield, E. C., & Soble, J. R. (2021). Assessment of differential neurocognitive performance based on the number of performance validity tests failures: A cross-validation study across multiple mixed clinical samples. *The Clinical Neuropsychologist*. https://doi.org/10.1080/13854046.2021.1900398

Keezer, R. D., Leib, S. I., Scimeca, L. M., Smith, J. T., Holbrook, L. R., Sharp, D. W., Jennette, K. J., Ovsiew, G. P., Resch, Z. J., & Soble, J. R. (2021). Masking effect of high IQ on the Rey Auditory Verbal Learning Test in an adult sample with attention deficit/hyperactivity disorder. *Applied Neuropsychology: Adult*. https://doi.org/10.1080/23279095.2021.1983575

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*(4), 666–679. https://doi.org/10.1080/13854040701494987

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*(4), 625–630.

Leib, S. I., Schieszler-Ockrassa, C., White, D. J., Gallagher, V. T., Carter, D. A., Basurto, K. S., Ovsiew, G. P., Resch, Z. J., Jennette, K. J., & Soble, J. R. (2021). Concordance between the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) and Clinical Assessment of Attention Deficit-Adult (CAT-A) over-reporting validity scales for detecting invalid ADHD symptom reporting. *Applied Neuropsychology: Adult*. https://doi.org/10.1080/23279095.2021.1894150

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., & Barker, J. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist, 24*, 1204–1237.

Martin, P. K., & Schroeder, R. W. (2020). Base rates of invalid test performance across clinical non-forensic contexts and settings. *Archives of Clinical Neuropsychology, 35*(6), 717–725. https://doi.org/10.1093/arclin/acaa017

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist, 29*(6), 741–776. https://doi.org/10.1080/13854046.2015.1087597

Musso, M. W., & Gouvier, W. D. (2014). "Why is this so hard?" A review of detection of malingered ADHD in college students. *Journal of Attention Disorders, 18*(3), 186–201.

Olsen, D. H., Schroeder, R. W., Heinrichs, R. J., & Martin, P. K. (2019). Examination of optimal embedded PVTs within the BVMT-R in an outpatient clinical sample. *The Clinical Neuropsychologist, 33*(4), 732–742.

Ovsiew, G. P., Resch, Z. J., Nayar, K., Williams, C. P., & Soble, J. R. (2020). Not so fast! Limitations of processing speed and working memory indices as embedded performance validity tests in a mixed neuropsychiatric sample. *Journal of Clinical and Experimental Neuropsychology, 42*, 473–484. https://doi.org/10.1080/13803395.2020.1758635

Pievsky, M. A., & McGrath, R. E. (2018). The neurocognitive profile of attention-deficit/hyperactivity disorder: A review of meta-analyses. *Archives of Clinical Neuropsychology, 33*(2), 143–157. https://doi.org/10.1093/arclin/acx055

Pliskin, J. I., DeDios-Stern, S., Resch, Z. J., Saladino, K. F., Ovsiew, G. P., Carter, D. A., & Soble, J. R. (2021). Comparing the psychometric properties of eight embedded performance validity tests in the rey auditory verbal learning test, wechsler memory scale logical memory, and brief visuospatial memory test–revised recognition trials for detecting invalid test performance. *Assessment, 28*(8), 1871–1881. https://doi.org/10.1177/1073191120929093

Poreh, A., Bezdicek, O., Korobkova, I., Levin, J. B., & Dines, P. (2016). The Rey Auditory Verbal Learning Test forced-choice recognition task: Base-rate data and norms. *Applied Neuropsychology: Adult, 23*, 155–161.

Poynter, K., Boone, K. B., Ermshar, A., Miora, D., Cottingham, M., Victor, T. L., Ziegler, E., Zeller, M. A., & Wright, M. (2019). Wait, there's a baby in this bath water! Update on quantitative and qualitative cutoffs for Rey 15-Item Recall and Recognition. *Archives of Clinical Neuropsychology, 34*, 1367–1380. https://doi.org/10.1093/arclin/acy087

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology, 31*(3), 206–230. https://doi.org/10.1093/arclin/acw007

Resch, Z. J., Pham, A. T., Abramson, D. A., White, D. J., DeDios-Stern, S., Ovsiew, G. P., Castillo, L. R., & Soble, J. R. (2020). Examining independent and combined accuracy of embedded performance validity tests in the California Verbal Learning Test-II and Brief Visuospatial Memory Test-Revised for detecting invalid performance. *Applied Neuropsychology Adult*. https://doi.org/10.1080/23279095.2020.1742718

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique [The psychological examination in cases of traumatic encephalopathy]. *Archives De Psychologie, 28*, 215–285.

Rhoads, T., Neale, A. C., Resch, Z. J., Cohen, C. D., Keezer, R. D., Cerny, B. M., Jennette, K. J., Ovsiew, G. P., & Soble, J. R. (2021). Psychometric implications of failure on one performance validity test: A cross-validation study to inform criterion group definition. *Journal of Clinical and Experimental Neuropsychology, 43*(5), 437–448. https://doi.org/10.1080/13803395.2021.1945540

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

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment, 19*(1), 21–30.

Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology, 35*(6), 735–764. https://doi.org/10.1093/arclin/acaa019

Soble, J. R., Alverson, W. A., Phillips, J. I., Critchfield, E. A., Fullen, C., O'Rourke, J. J. F., Messerly, J., Highsmith, J. M., Bailey, K. C., Webber, T. A., & Marceaux, J. M. (2020). Strength in numbers or quality over quantity? Examining the importance of criterion measure selection to define validity groups in performance validity test (PVT) research. *Psychological Injury and Law, 13*, 44–56. https://doi.org/10.1007/s12207-019-09370-w

Soble, J. R., Sharp, D. W., Carter, D. A., Jennette, K. J., Resch, Z. J., Ovsiew, G. P., & Critchfield, E. A. (2021). Cross-validation of a forced-choice validity indicator to enhance the clinical utility of the Rey Auditory Verbal Learning Test. *Psychological Assessment, 33*(6), 568–573.

Suhr, J., Hammers, D., Dobbins-Buckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Archives of Clinical Neuropsychology, 23*, 521–530.

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology, 14*, 189–207.

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Participants, C. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologis*. https://doi.org/10.1080/13854046.2021.1896036

Teter, C. J., McCabe, S. E., Cranford, J. A., Boyd, C. J., & Guthrie, S. K. (2005). Prevalence and motive for illicit use of prescription stimulants in an undergraduate student simple. *Journal of American College Health, 53*(6), 253–262. https://doi.org/10.1177/1073191118804874

Webber, T. A., Critchfield, E. A., & Soble, J. R. (2020). Convergent, discriminant, and concurrent validity of non-memory-based performance validity tests. *Assessment, 27*(7), 1399–1415. https://doi.org/10.1177/1073191118804874

Webber, T. A., & Soble, J. R. (2018). Utility of various WAIS-IV Digit Span indices for identifying noncredible performance validity among cognitively impaired and unimpaired examinees. *The Clinical Neuropsychologist, 32*(4), 657–670. https://doi.org/10.1080/13854046.2017.1415374

White, D. J., Korinek, D., Bernstein, M. T., Ovsiew, G. P., Resch, Z. J., & Soble, J. R. (2020a). Cross-validation of non-memory-based embedded performance validity tests for detecting invalid performance among patients with and without neurocognitive impairment. *Journal of Clinical and Experimental Neuropsychology, 42*(5), 459–472. https://doi.org/10.1080/13803395.2020.1758634

White, D. J., Ovsiew, G. P., Rhoads, T., Resch, Z. J., Lee, M., Oh, A. J., & Soble, J. R. (2020b). The divergent roles of symptom and performance validity in the assessment of ADHD. *Journal of Attention Disorders*. https://doi.org/10.1177/1087054720964575

Whitney, K. A., & Davis, J. J. (2015). The non-credible score of the Rey Auditory Verbal Learning Test: Is it better at predicting non-credible neuropsychological test performance than the RAVLT recognition score? *Archives of Clinical Neuropsychology, 30*, 130–138.