Article

Journal of Attention Disorders
2023, Vol. 27(1) 80–88

DOI: 10.1177/10870547221121781
journals.sagepub.com/home/jad
SAGE

Settling the Score: Can CPT-3 Embedded Validity Indicators Distinguish Between Credible and Non-Credible Responders Referred for ADHD and/or SLD?

Anthony Robinson[1], Christopher Reed[1], Katrail Davis[1], Ross Divers[1], Luke Miller[1], Laszlo A. Erdodi[2], and Matthew Calamia[1]

Abstract

Objective: The purpose of the present study was to further investigate the clinical utility of individual and composite indicators within the CPT-3 as embedded validity indicators (EVIs) given the discrepant findings of previous investigations. **Methods:** A total of 201 adults undergoing psychoeducational evaluation for ADHD and/or Specific Learning Disorder (SLD) were divided into credible (n = 159) and non-credible (n = 42) groups based on five criterion measures. **Results:** Receiver operating characteristic curves (ROC) revealed that 5/9 individual indicators and 2/4 composite indicators met minimally acceptable classification accuracy of ≥0.70 (AUC = 0.43–0.78). Individual (0.16–0.45) and composite indicators (0.23–0.35) demonstrated low sensitivity when using cutoffs that maintained specificity ≥90%. **Conclusion:** Given the lack of stability across studies, further research is needed before recommending any specific cutoff be used in clinical practice with individuals seeking psychoeducational assessment. (*J. of Att. Dis. 2023; 27(1) 80-88*)

Keywords

CPT, performance validity, embedded validity, ADHD, SLD

Neuropsychologists' support and implementation of validity testing has been consistently high with the overwhelming majority using performance (PVTs) and symptom (SVTs) validity tests in their clinical and forensic practice (Sweet et al., 2015, 2021). PVTs are used to flag underperformance on neuropsychological assessments and can help clinicians avoid misdiagnosis and unneeded additional testing (Lippa, 2018). PVTs as well as SVTs are also important in the context of psychoeducational evaluations given that they are typically sought to gain access to academic accommodations such as extra time for testing and/or stimulant medications (Lovett & Harrison, 2021; Suhr & Berry, 2017). Traditionally, standalone PVTs have been the standard to use in evaluations, however embedded validity indices (EVIs) such as those derived from Continuous Performance Tests have received increased attention as they are more efficient and less susceptible to coaching (Miele et al., 2012). The present study focuses on the use of Continuous Performance Tests as EVI's in psychoeducational evaluations.

Continuous Performance Tests were first developed to measure lapses in sustained attention in those with brain damage and have since been used in the assessment of other disorders including ADHD (Albrecht et al., 2015). Continuous Performance Tests such as the Conners Continuous Performance Test (CPT-II; Conners, 2000) have been subject to extensive research on the utility of individual and composite indicators as EVIs (e.g., Erdodi et al., 2014, 2017, 2018; J. S. Ord et al., 2010). These studies largely support the use of several individual and composite indicators from the CPT-II as EVIs in those with a traumatic brain injury (TBI; Lange et al., 2013) and in mixed clinical samples (e.g., Sharland et al., 2018). For example, the use of CPT-II Error indices such as OMI (Busse & Whiteside, 2012; J. S. Ord et al., 2010; Sharland et al., 2018), COM (Lange et al., 2013; Sharland et al., 2018), and Total errors (Busse & Whiteside, 2012; Sharland et al., 2018; Shura et al., 2016) has been supported. With regard to composite indicators, studies have demonstrated that aggregating CPT-II indices often improves classification accuracy (Erdodi et al., 2014, 2018).

[1]Louisiana State University, Baton Rouge, USA
[2]University of Windsor, Ontario, Canada

Corresponding Author:
Matthew Calamia, Department of Psychology, Louisiana State University, 33 Johnston Hall, Baton Rouge, LA 70803, USA.
Email: mcalamia@lsu.edu

Less research, however, has been conducted on the newest version of the CPT, the CPT-3 (Conners, 2014) which differs from its predecessor in how it was standardized, the types of scores it yields, and the algorithms used to derive them (Conners, 2014). For example, the CPT-3 was normed utilizing a sample more representative of the US population than the CPT-II normative sample, Hit Reaction Time Standard Error score was replaced with HRT SD, and HRT Standard Error Block Change and HRT Standard Error ISI Change were not included in the CPT-3. Changes were also made to the overall testing and reporting format, with the removal of raw scores from the score report (Conners, 2014). Of note, Hit Reaction Time Standard Error was included in several composite EVIs and a number of EVIs have been derived using raw scores from the CPT-II. Given the changes to the normative sample as well as scoring changes there is a need to examine existing CPT-II EVIs to determine if they are compatible with the CPT-3. Further, given that CPT-II EVIs have primarily been examined in the context of TBI there is a need to cross-validate these findings in other samples, including ADHD and psychoeducational samples more broadly.

To date, only two studies have examined EVIs within the CPT-3. First, A. S. Ord et al. (2021) used a sample of 197 veterans referred for ADHD. All participants were given the Test of Memory Malingering Trial 1 (TOMM1; Tombaugh, 1996) and the CPT-3. Participants scoring <42 on TOMM1 were considered non-credible. The authors found that several indicators including Omissions (OMI), Commissions (COM), Detectability (DET), Hit Reaction Time Standard Deviation (HRT SD), and HRT Inter-Stimulus Interval (ISI) Change significantly predicted validity status. Of note, HRT SD and HRT ISI Change demonstrated the highest area under the curve (AUC) values (0.72 and 0.71, respectively) but demonstrated low sensitivity when maximizing specificity. In addition, two composite indicators derived by summing the T scores of individual CPT-3 indicators (i.e., CEVI 1 and CEVI 3) met minimal acceptable classification accuracy and modest increases in sensitivity when maximizing sensitivity (A. S. Ord et al., 2021). Overall, the results suggest that composite indicators of CPT-3 increase classification accuracy as has been shown with the CPT-II.

Scimeca et al. (2021) further examined the utility of the CPT-3 as an EVI in a sample of 201 civilian adults referred for ADHD. Participants who failed 2+ PVTs out of 7 possible criterion measures were included in the non-credible group. The authors found that six indicators including COM, HRT SD, HRT ISI Change, OMI, DET, and Variability (VAR) significantly predicted validity status. Of note, only OMI and VAR met minimal acceptable classification accuracy ($AUC=0.70$). The authors also examined several composite scores including two derived from the CPT-II (CVI-5A, CVI-5B; Erdodi et al., 2014) which aggregate the number of *T*-scores at or above a specified cutoff for individual CPT-II indicators, the top two performing composites from A. S. Ord et al. (2021; i.e., CEVI 1 and CEVI 3), and two newly created composites, one of which includes aggregating the number of *T*-scores at or above a specified cutoff (Composite 1) and the other includes summing *T*-Scores (Composite 2). Although the six composite indicators significantly predicted validity status, they did not provide better classification accuracy than the individual indicators (Scimeca et al., 2021). Further, indicators that were found to distinguish between credible and non-credible responders yielded low sensitivity (0.13–0.38) and were discrepant from those found in A. S. Ord et al. (2021). As a result, the authors concluded that EVI's derived from the CPT-3 lack stability and may not be appropriate for individuals referred for ADHD (Scimeca et al., 2021).

Given the discrepancies regarding the utility of composite indicators as well as individual CPT-3 indicators as EVIs including VAR, OMI, HRT SD, and HRT ISI Change, the current study seeks to further investigate the effectiveness of both individual and composite scores from the CPT-3 for detecting non-credible performance in sample of adults referred for psychoeducational evaluation. Similarly to Scimeca et al. (2021), several freestanding and embedded PVTs were used to create the credible and non-credible groups. Given the similar criterion grouping approach and similarity between samples, we hypothesized that composite EVIs would not provide better classification accuracy than the individual indicators.

Method

Participants

Participants were 307 individuals referred to a university-affiliated clinic for a psychoeducational evaluation for concerns related to ADHD, SLD, or both. Given the focus of the present study, individuals were excluded if they did not receive the CPT-3 ($n=33$) or if individuals were administered less than two freestanding or embedded PVTs ($n=2$). Five independent criterion PVTs were used to determine validity status (see Table 1). Consistent with current practice guidelines, those who failed 0 PVTs were included in the credible group and those who failed ≥ 2 PVTs were included in the non-credible group (Sherman et al., 2020). Because the sample included a mix of individuals with identifiable external incentive to appear impaired, those who failed 1 PVT ($n=71$) were excluded from analyses (Schroeder, 2018). The final sample included 201 participants ($n=159$ credible group, $n=42$ non-credible group). See Table 2 for demographic characteristics of the final sample.

Procedures

A comprehensive psychoeducational assessment was completed as part of clinical services and was not part of a

Table 1. Criterion Performance Validity Tests in Full Sample.[a]

Performance validity test	Cut-score(s)	Reference	Sample failure rate
Dot counting test	E-score ≥ 14	Boone et al. (2002)	29/221 (13.1%)
VSVT	Difficult score < 19	Frazier et al. (2008)	45/237 (19.0%)
b Test	*E*-score > 120	Marshall et al. (2010)	32/259 (12.4%)
Green's MSVT	IR, DR, or CNS ≤ 85	Green (2004)	26/206 (12.6%)
WAIS-IV digit span	Reliable digit span ≤ 7	Marshall et al. (2010)	39/254 (15.4%)

Note. DCT = dot counting test; VSVT = Victoria symptom validity test; MSVT = medical symptom validity test; WAIS-IV = Wechsler Adult Intelligence Scale-Fourth Edition.
[a]Includes those who failed 1 PVT only ($n = 272$).

Table 2. Demographic Characteristics of Final Sample.

	Credible ($n = 159$)	Non-credible ($n = 42$)		
	M (*SD*)	*M* (*SD*)	t/χ^2	Cohen's *d*
Age	23.04 (6.80)	22.93 (4.38)	0.10	
Years of education	13.94 (1.99)	14.29 (2.10)	−0.99	
Sex			5.66	
Male	81 (86.2%)	13 (13.8%)		
Female	76 (72.4%)	29 (27.6%)		
Race			6.46	
White	127 (81.4%)	29 (18.6%)		
Black	18 (66.7%)	9 (33.3%)		
Hispanic	38 (23%)	8 (25%)		
Asian	9 (66.7%)	3 (33.3%)		
Other	5 (83.3%)	1 (16.7%)		
Referral			3.71	
ADHD	109 (82.6%)	23 (17.4%)		
SLD	32 (76.2%)	10 (23.8%)		
SLD + ADHD	18 (66.7%)	9 (33.3%)		
Student status			2.42	
High school	5 (83.3%)	1 (16.7%)		
Non-student/community member	11 (64.7%)	6 (35.3%)		
Undergraduate student	121 (80.7%)	29 (19.3%)		
Graduate student	22 (78.6%)	6 (21.4%)		
Accommodations status			3.21	
Seeking accommodations	105 (75.5%)	34 (24.5%)		
Not seeking accommodations	38 (88.4%)	5 (11.6%)		
CPT-3 score				
Omissions	47.82 (9.49)	55.74 (13.83)	−3.50**	0.66
Commissions	53.60 (12.07)	62.31 (11.40)	−4.20***	0.74
Variability	48.69 (11.34)	61.05 (14.48)	−5.13***	0.95
Detectability	51.28 (11.42)	61.93 (11.40)	−5.38***	0.93
Perseveration	51.07 (11.96)	55.86 (13.34)	−2.25*	0.37
HRT	49.39 (10.57)	52.62 (11.77)	−1.72***	0.28
HRT *SD*	49.86 (12.23)	60.64 (14.81)	−4.86***	0.79
HRT block Δ	49.73 (11.55)	48.55 (12.72)	0.58	0.09
HRT ISI Δ	52.96 (12.36)	55.57 (12.54)	−1.21	0.20

Note. *M* = mean; *SD* = standard deviation; SLD = specific learning disorder; CPT-3 = Conners continuous performance task third edition; HRT = hit reaction time; Δ = change; ISI = interstimulus interval.
Aadjusted *p*-values: *$p < .05$. **$p < .01$. ***$p < .001$.

research battery. Therefore, the order and specific tests included varied somewhat based on the clinical needs of the individual. However, a core battery of tests was consistently administered as part of the psychoeducational assessment in accordance with the standardized instructions contained in the respective test manual. All assessments were completed by supervised doctoral students and included clinical interviews, cognitive and academic measures as well as several freestanding and embedded PVTs. Following assessment, tests were scored, and a report was completed, including a final diagnosis if warranted. All research procedures were approved by Louisiana State University's Institutional Review Board.

Measures

Conners Continuous Performance Test-3 (CPT-3)

The CPT-3 is computerized cognitive test of sustained attention, impulsivity, inattention, and vigilance which requires individuals to press the spacebar for all letters when they appear, except for the letter "X" (Conners, 2014). The CPT-3 generates multiple T scores describing the individual's response style, ability to discriminate targets from non-targets, error types, and reaction time statistics and yields (Conners, 2014). Specifically, the CPT-3 yields the following nine scores: detectability (DET), omission errors (OMI), commission errors (COM), perseverations (PER), hit reaction time (HRT), HRT Standard Deviation (*SD*), variability (VAR), HRT block change, and HRT inter-stimulus interval change (HRT ISI Change). The CPT-3 demonstrates good internal consistency (spilt-half reliability = .92 and .94, for the normative sample and clinical sample, respectively; Conners, 2014).

Analytical Plan

Correlations were conducted to assess the relationship between the criterion PVTs and CPT-3 scores among credible responders. Independent *t*-tests were conducted to examine differences in performance across all CPT-3 scores between the credible and non-credible groups. Next, receiver operating characteristic (ROC) curve analyses were then performed to test the accuracy of the individual CPT-3 scores for detecting invalid performance. Another set of ROC curve analyses were then conducted to determine if proposed composite CPT-3 scores improved classification accuracy. Two CEVIs identified by Ord et al. (2021) and two CEVIs identified by Scimeca et al. (2021) were examined (see Table 4 for description of composites). Minimal acceptable classification accuracy was set at an area under the curve (AUC) of ≥0.70 for all ROC curve analyses (Hosmer et al., 2013). Cutoffs that maximized sensitivity while maintaining ≥90% specificity were selected as optimal cut-scores for all significant AUCs (Boone, 2013) that met minimal acceptable classification accuracy. A correction for multiple comparisons was applied to each set of analyses utilizing the FDR method (Benjamini & Hochberg, 1995).

Results

Descriptives and correlational analyses can be found in Tables 1 and 3, respectively. Correlational analyses revealed non-significant to small correlations ($r = <.01$–.28) between the criterion PVTs and CPT-3 scores. After correction for multiple comparisons significant group differences were found for DET, HRT SD, VAR, COM, HRT, and OMI with the non-credible group performing worse than the credible group ($d = 0.28$–0.95). VAR and DET yield the largest differences between the groups ($d = 0.95$ and 0.93, respectively).

Individual CPT-3 Indicators

ROC analyses with the individual indicators revealed that VAR, DET, HRT SD, OMI, COM, and PER had significant AUCs even after correction for multiple comparisons (Table 4). Of note, all of these indicators had an AUC of ≥0.70 with the exception of PER (AUCs = 0.70–0.77). When maintaining specificity at ≥0.90 sensitivity was low for all individual indicators, ranging from 0.16 to 0.45. Of note, DET yielded the highest sensitivity at a cut score of $T \geq 63$ (sensitivity = 0.45, specificity = 0.90).

Composite CPT-3 Indicators

All ROC curve analyses of composite scores had significant AUC's even after correction for multiple comparisons (Table 4). However, only the Ord CEVI 2 and Scimeca CEVI 2 composite scores were above the minimum AUC (AUC = 0.76 and 0.78, respectively). When maintaining specificity at ≥0.90 all four composite scores yielded low sensitivity (0.19–0.35), with the Scimeca CEVI 1 composite yielding the highest sensitivity at a cut score of ≥ 2 (sensitivity = 0.35, specificity = 0.91).

Discussion

The present study further investigated the use of EVI indicators from the CPT-3 for detecting non-credible performance in sample of adults referred for psychoeducational evaluation given the discrepant findings across studies. Several freestanding and embedded criterion PVT's were used to establish credible and non-credible groups. Base rates of failure on criterion measures prior to removing

Table 3. Correlations Between CPT-3 Scores and Criterion Performance Validity Tests.

	VSVT	bT-E	IR	DR	CNS	RDS	DET	OM	COM	PER	HRT	HRT SD	VAR	HRT Bk Δ	HRT ISI Δ
DCT-E	0.01	−0.03	0.02	0.04	−0.01	−0.09	0.24**	0.28**	0.18	0.22*	0.11	0.23**	0.17	0.07	0.07
VSVT	—	0.06	0.09	0.02	−0.03	0.1	−0.21*	−0.09	−0.26	−0.03	0.09	−0.05	−0.08	−0.06	−0.07
bT-E		—	−0.03	−0.09	0.02	−0.05	0.01	−0.03	0.002	0.02	0.07	0.04	0.06	0.05	0.02
IR			—	0.03	0.59**	0.01	0.03	0.03	−0.01	0.07	0.08	0.06	0.04	−0.05	0.05
DR				—	0.07	0.05	0.02	−0.03	0.000	−0.05	−0.07	−0.08	−0.06	0.09	−0.07
CNS					—	0.01	−0.04	−0.06	−0.06	−0.04	0.02	0.004	0.003	0.01	0.01
RDS						—	−0.08	−0.05	−0.13	0.02	0.02	0.04	0.05	−0.04	0.17*
DET							—	0.76**	0.90**	0.67**	0.05	0.54**	0.60**	0.30**	0.33*
OM								—	0.51**	0.56**	0.44**	0.63**	0.62**	0.33**	0.43**
COM									—	0.53**	−0.12	0.38**	0.50**	0.25**	0.24**
PER										—	0.20*	0.54**	0.62**	0.40**	0.38**

Note. DIF = Victoria symptom validity test difficult score; bT-E = b Test E-score; IR = Green's medical symptom validity test immediate recall; DR = Green's medical symptom validity test delayed recall; CNS = Green's medical symptom validity test consistency; RDS = Wechsler Adult Intelligence Scale-fourth edition (WAIS-IV) reliable digit span forward and backward; CPT-3 = Conner's continuous performance task third edition; OM = CPT-3 omissions; COM = CPT-3 commissions; VAR = CPT-3 variability; DET = CPT-3 detectability; PER = CPT-3 perseverations; HRT = CPT-3 hit reaction time; HRT SD = CPT-3 hit reaction time standard deviation; HRT Bk Δ = CPT-3 hit reaction time interstimulus interval block change; HRT ISI Δ = CPT-3 hit reaction time interstimulus interval change.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$

individuals who failed 1 PVT ranged from 12.4% to 19.0% and is in line with rates of failure found in other psychoeducational samples (Harrison et al., 2021). The non-credible group was worse at discriminating between targets and non-targets, making more commission errors and a higher number of random anticipatory responses than the credible group on the CPT-3. The non-credible group also had more inconsistencies in response speed and greater variability in reaction time and responded faster to stimuli. However, similarly to previous investigations, scores across both groups were still largely within the normal range.

Several individual CPT-3 indicators had significant AUC's that met minimal classification accuracy including OMI (0.35 SN/0.91 SP), COM (0.33 SN/0.91 SP), VAR (0.38 SN/0.90 SP), DET (0.45 SN/0.90 SP), and HRT SD (0.26 SN/0.91 SP). AUC's for individual indicators were nearly identical to Scimeca et al. (2021) for VAR (AUC 0.77 vs. 0.70) and OMI (AUC 0.72 vs. 0.70) and was identical to A. S. Ord et al. (2021) for HRT SD (AUC = 0.72). Interestingly, DET and COM met minimal classification accuracy in the current study but did not meet minimal classification accuracy in either of the previous investigations with the CPT-3 (A. S. Ord et al., 2021; Scimeca et al., 2021). However, investigations with the CPT-II have demonstrated that COM can accurately distinguish between credible and non-credible responders (e.g., Lange et al., 2013; Sharland et al., 2018). With regard to PPV and NPV, at optimal cutoffs each indicator demonstrated good to acceptable NPV (0.88–0.90) at lower base rates similarly to previous CPT-3 studies (NPV = 0.86–0.88, A. S. Ord et al., 2021; NPV = 0.91–0.92, Scimeca et al., 2021). However, PPV was very low (0.27–0.44), suggesting that at lower base rates of failure optimal cutoffs may be associated with a high likelihood of falsely identifying an individual as non-credible given failure.

Of note, there were no individual CPT-3 indicators that performed well across all three CPT-3 studies, suggesting an overall lack of stability. Alternatively, methodological differences (e.g., sample characteristics, number of PVTs employed, and type of PVTs employed) may account for discrepancies across studies. However, it is important to note that despite these methodological differences, PER, HRT, and HRT Block Change were found to perform poorly across all three studies. This is in contrast with studies on the CPT-II which found that PER (Erdodi et al., 2014, 2018) and HRT (J. S. Ord et al., 2010) demonstrated acceptable classification accuracy. However, these studies examined the utility of these indicators in TBI samples. This suggests that while these indicators may be appropriate for use in TBI they may not be appropriate for determining validity status in ADHD or combined ADHD/SLD assessments.

Regarding composite scores, only ORD CEVI 2 (CEVI 3; A. S. Ord et al., 2021; 0.26 SN/0.91 SP) and Scimeca CEVI 2 (Composite 2; Scimeca et al., 2021; 0.26 SN/0.90) met minimal classification accuracy but demonstrated low sensitivity when maintaining specificity ≥0.90. Although Scimeca CEVI 1 did not meet minimal classification accuracy it did yield the highest sensitivity (0.35) out of all the composite scores at a cut score of ≥2 and demonstrates sensitivity comparable to other EVIs used in ADHD evaluations such as reliable digit span (Bing-Canar et al., 2022), the Trail Making Test (White et al., 2020), and Verbal Fluency Test (White et al., 2020). Moreover, CPT-3

Table 4. CPT-3 Indicator Accuracy for Detecting Invalidity.

CPT-3 indicators	AUC	Cutoff	Non-CR	CR	EVI Fail	EVI False	Sen.	Spec.	BR 15% PPV	BR 15% NPV	BR 30% PPV	BR 30% NPV	BR 50% PPV	BR 50% NPV
Omissions	0.72	≥54	16	141	26	18	0.38	0.89	0.37	0.89	0.59	0.76	0.77	0.58
		≥55	16	141	26	18	0.38	0.89	0.37	0.89	0.59	0.76	0.77	0.58
		≥56	**15**	**145**	**27**	**14**	**0.36**	**0.91**	**0.41**	**0.88**	**0.63**	**0.76**	**0.80**	**0.58**
		≥57	13	147	29	12	0.31	0.92	0.41	0.88	0.63	0.75	0.80	0.57
		≥58	13	147	29	12	0.31	0.92	0.41	0.88	0.63	0.75	0.80	0.57
		≥59	12	149	30	10	0.29	0.94	0.44	0.88	0.66	0.75	0.81	0.56
Commissions	0.70	≥67	16	140	26	19	0.38	0.88	0.36	0.88	0.57	0.76	0.76	0.58
		≥68	**15**	**143**	**27**	**16**	**0.36**	**0.90**	**0.38**	**0.88**	**0.60**	**0.76**	**0.78**	**0.58**
		≥69	14	146	28	13	0.33	0.92	0.41	0.88	0.63	0.76	0.80	0.57
		≥70	13	149	29	10	0.31	0.94	0.46	0.88	0.67	0.76	0.83	0.57
Variability	0.78	≥59	20	139	22	19	0.48	0.88	0.41	0.90	0.62	0.79	0.79	0.62
		≥60	**16**	**143**	**26**	**15**	**0.38**	**0.91**	**0.41**	**0.89**	**0.63**	**0.77**	**0.80**	**0.59**
		≥61	14	144	28	14	0.33	0.91	0.39	0.88	0.61	0.75	0.78	0.57
		≥62	14	144	28	14	0.33	0.91	0.39	0.88	0.61	0.75	0.78	0.57
		≥63	12	145	30	13	0.29	0.92	0.38	0.87	0.59	0.74	0.77	0.56
		≥64	12	145	30	13	0.29	0.92	0.38	0.87	0.59	0.74	0.77	0.56
		≥65	12	146	30	12	0.29	0.92	0.39	0.88	0.61	0.75	0.79	0.56
Detectability	0.76	≥62	22	131	20	28	0.52	0.82	0.34	0.90	0.56	0.80	0.74	0.63
		≥63	**18**	**137**	**22**	**15**	**0.45**	**0.90**	**0.44**	**0.90**	**0.66**	**0.79**	**0.82**	**0.62**
		≥64	18	144	24	15	0.43	0.91	0.44	0.89	0.66	0.78	0.81	0.61
		≥65	16	146	26	13	0.38	0.92	0.45	0.89	0.66	0.77	0.82	0.59
Hit reaction time *SD*	0.73	≥65	**12**	**143**	**30**	**16**	**0.29**	**0.90**	**0.33**	**0.87**	**0.54**	**0.74**	**0.73**	**0.55**
		≥66	11	145	31	14	0.26	0.91	0.34	0.87	0.56	0.74	0.74	0.55
		≥67	10	145	32	14	0.24	0.91	0.32	0.87	0.53	0.73	0.73	0.54
Perseverations	0.64	≥64	9	137	33	21	0.21	0.87	0.22	0.86	0.40	0.72	0.61	0.52
		≥65	**7**	**146**	**35**	**12**	**0.17**	**0.92**	**0.27**	**0.86**	**0.48**	**0.72**	**0.68**	**0.52**
		≥ 70	7	146	35	12	0.17	0.92	0.27	0.86	0.48	0.72	0.68	0.52
		≥71	6	147	36	11	0.14	0.93	0.26	0.82	0.46	0.71	0.67	0.52
Ord 1	0.68	≥124	10	143	32	16	0.24	0.90	0.29	0.86	0.50	0.73	0.70	0.54
		≥128	**10**	**144**	**32**	**15**	**0.24**	**0.91**	**0.30**	**0.87**	**0.51**	**0.73**	**0.71**	**0.54**
		≥130	9	144	33	15	0.21	0.91	0.28	0.86	0.49	0.72	0.69	0.53
		≥132	8	145	34	14	0.19	0.91	0.27	0.86	0.48	0.72	0.68	0.52
Ord 2	0.76	**≥240**	**12**	**143**	**30**	**16**	**0.29**	**0.90**	**0.33**	**0.87**	**0.54**	**0.74**	**0.73**	**0.55**
		≥246	11	144	31	15	0.26	0.91	0.36	0.87	0.57	0.74	0.76	0.55
		≥248	11	144	31	15	0.26	0.91	0.36	0.87	0.57	0.74	0.76	0.55
		≥250	11	146	31	13	0.26	0.92	0.39	0.87	0.61	0.74	0.78	0.55
		≥254	9	143	33	11	0.21	0.93	0.34	0.87	0.56	0.73	0.75	0.54
Scimeca 1	0.70	≥1	25	121	17	37	0.60	0.77	0.30	0.91	0.52	0.81	0.71	0.65
		≥2	**15**	**144**	**27**	**14**	**0.36**	**0.91**	**0.48**	**0.90**	**0.69**	**0.80**	**0.84**	**0.63**
		≥3	8	150	34	8	0.19	0.95	0.38	0.86	0.61	0.73	0.79	0.53
Scimeca 2	0.79	≥300	13	141	29	17	0.31	0.89	0.33	0.87	0.55	0.75	0.74	0.56
		≥302	**12**	**142**	**30**	**16**	**0.29**	**0.90**	**0.33**	**0.87**	**0.54**	**0.74**	**0.73**	**0.55**
		≥305	11	143	31	15	0.26	0.91	0.32	0.87	0.54	0.74	0.73	0.55
		≥310	10	145	32	13	0.24	0.92	0.33	0.87	0.55	0.73	0.74	0.54

Note. Optimal cutoffs in bold. CPT-3 = continuous performance task third edition; HRT SD = hit reaction time standard deviation; AUC = area under curve; Non-CR = non-credible responders; CR = credible responders; EVI Fail = embedded validity indicator failed to identify; EVI False = embedded validity indicator falsely identified; Sen = sensitivity; Spec = specificity; BR = base rate. Ord 1 = HRT SD + HRT ISI Change; Ord 2 = OMI + COM + HRT SD + HRT ISI Change; Scimeca 1 = Total number of *T*-scores above cutoff using Omissions ≥58, Commissions ≥68, Hit RT SD ≥67, VAR ≥65, Hit ISI Change ≥72. Scimeca 2 = OMI + COM + Hit RT SD + VAR + Hit ISI Change.

composite indicators overall did not improve classification accuracy, which is consistent with Scimeca et al. (2021). Studies examining the utility of composite indicators with the CPT-II however, have demonstrated improved accuracy of composite variables over individual indicators in TBI (e.g., Erdodi et al., 2014, 2017). Discrepancies between findings for the CPT-3 and CPT-II may be related to methodological differences (e.g., sample size, population, and validity grouping methods) and/or differences in test properties (e.g., normative ranges; Scimeca et al., 2021).

The present study found some support for previously investigated individual and composite EVIs within the CPT-3. However, it is important to note that the cut scores derived in the present study were much lower than previous investigations. For example, the present study found that a cut score of ≥ 240 on ORD CEVI 2 demonstrates the greatest sensitivity while maximizing specificity whereas A. S. Ord et al. (2021) found a cut score of >237 and Scimeca et al. (2021) identified a cut score of ≥256. Taken together, these findings suggest that although select individual and composite CPT-3 variables may accurately differentiate credible responding from non-credible performance, they may not be appropriate for use in those referred for psychoeducational evaluation given their overall lack of stability. As such clinicians should be cautious when interpreting EVIs on the CPT-3 with individuals presenting for ADHD or SLD at present. Additional studies are needed to further examine and cross-validate established EVIs and cutoff scores before implementing their use into clinical practice. Alternative measures such as the MOXO-d-CPT may be of utility as it has demonstrated high specificity and good sensitivity for detecting feigned ADHD. For example, simulation studies have demonstrated 91% to 96% specificity and 76% to 89% sensitivity (Berger et al., 2021; Winter & Braw, 2022). However, research using a clinical sample known-groups design has not been conducted.

Although the present study had a number of strengths including a mixed sample of individuals referred for ADHD and/or specific learning disability and utilized multiple embedded and free standing PVTs, there are several notable limitations. First, the size of the non-credible group was small, though comparable to previous investigations. Future studies should include larger samples given that cutoff scores derived from small sample sizes can be less stable than cut-off scores derived from large sample sizes (Greve & Bianchini, 2004). Secondly, this study was unable to perform subgroup analyses of ADHD and specific learning disability because there were not enough individuals in each group for analysis. Results may be different between these groups, given that those feigning specific learning disability may use different strategies than those who feign ADHD (e.g., Harrison et al., 2008, 2010). Therefore, future studies should further investigate the use of the CPT-3 as an embedded validity measure in ADHD and SLD separately.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Anthony Robinson https://orcid.org/0000-0003-1817-559X
Laszlo A. Erdodi https://orcid.org/0000-0003-0575-9991

References

Albrecht, B., Uebel-von Sandersleben, H., Wiedmann, K., & Rothenberger, A. (2015). ADHD history of the concept: The case of the continuous performance test. *Current Developmental Disorders Reports*, *2*, 10–22 https://doi.org/10.1007/s40474-014-0035-1

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berger, C., Lev, A., Braw, Y., Elbaum, T., Wagner, M., & Rassovsky, Y. (2021). Detection of feigned ADHD using the MOXO-d-CPT. *Journal of Attention Disorders*, *25*(7), 1032–1047. https://doi.org/10.1177/1087054719864656

Bing-Canar, H., Phillips, M. S., Shields, A. N., Ogram Buckley, C. M., Chang, F., Khan, H., Skymba, H. V., Ovsiew, G. P., Resch, Z. J., Jennette, K. J., & Soble, J. R. (2022). Cross-validation of multiple WAIS-IV Digit Span embedded performance validity indices among a large sample of adult Attention Deficit/Hyperactivity Disorder clinical referrals. *Journal of Psychoeducational Assessment*, *40*, 678–688. https://doi.org/10.1177/07342829221081921

Boone, K. B. (2013). Clinical practice of forensic neuropsychology. Guilford Press.

Boone, K. B., Lu, P., & Herzberg, D. (2002). The Dot Counting Test manual. Los Angeles, CA: Western Psychological Services.

Busse, M., & Whiteside, D. (2012). Detecting suboptimal cognitive effort: Classification accuracy of the Conner's continuous performance test-II, brief test of attention, and trail making test. *The Clinical Neuropsychologist*, *26*(4), 675–687. https://doi.org/10.1080/13854046.2012.679623

Conners, C. K. (2004). *Conners' continuous performance test (CPT-II): Technical guide and software manual*. Multi-Health Systems.

Conners, C. K. (2014). *Conners' continuous performance test – 3rd edition (CPT-3) manual*. Multi-Health Systems.

Erdodi, L. A., Abeare, C. A., Lichtenstein, J. D., Tyson, B. T., Kucharski, B., Zuccato, B. G., & Roth, R. M. (2017). Wechsler Adult Intelligence Scale-fourth edition (WAIS-IV) processing speed scores as measures of noncredible responding: The third generation of embedded performance validity indicators. *Psychological Assessment*, *29*(2), 148–157. https://doi.org/10.1037/pas0000319

Erdodi, L. A., Pelletier, C. L., & Roth, R. M. (2018). Elevations on select Conners' CPT-II scales indicate noncredible responding in adults with traumatic brain injury. *Applied Neuropsychology: Adult*, *25*(1), 19–28. https://doi.org/10.1080/23279095.2016.1232262

Erdodi, L. A., Roth, R. M., Kirsch, N. L., Lajiness-O'Neill, R., & Medoff, B. (2014). Aggregating validity indicators embedded in conners' CPT-II outperforms individual cutoffs at separating valid from invalid performance in adults with traumatic brain injury. *Archives of Clinical Neuropsychology*, *29*(5), 456–466. https://doi.org/10.1093/arclin/acu026

Frazier, T. W., Frazier, A. R., Busch, R. M., Kerwood, M. A., & Demaree, H. A. (2008). Detection of simulated ADHD and reading disorder using symptom validity measures. *Archives of Clinical Neuropsychology*, *23*(5), 501–509. https://doi.org/10.1016/j.acn.2008.04.001

Green, P. (2004). Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows: User's manual. Edmonton, Canada: Green's Publishing.

Greve, K. W., & Bianchini, K. J. (2004). Setting empirical cut-offs on psychometric indicators of negative response bias: A methodological commentary with recommendations. *Archives of Clinical Neuropsychology*, *19*, 533–541.

Harrison, A. G., Beal, A. L., & Armstrong, I. T. (2021). Predictive value of performance validity testing and symptom validity testing in psychoeducational assessment. *Applied Neuropsychology Adult*. Advance online publication. https://doi.org/10.1080/23279095.2021.1943396

Harrison, A. G., Edwards, M. J., Armstrong, I., & Parker, K. C. (2010). An investigation of methods to detect feigned reading disabilities. *Archives of Clinical Neuropsychology*, *25*(2), 89–98.

Harrison, A. G., Edwards, M. J., & Parker, K. C. (2008). Identifying students feigning dyslexia: Preliminary findings and strategies for detection. *Dyslexia*, *14*(3), 228–246.

Hosmer, D. W. J., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. New York: John Wiley & Sons.

Lange, R. T., Iverson, G. L., Brickell, T. A., Staver, T., Pancholi, S., Bhagwat, A., & French, L. M. (2013). Clinical utility of the Conners' continuous performance test-II to detect poor effort in U.S. Military personnel following traumatic brain injury. *Psychological Assessment*, *25*(2), 339–352. https://doi.org/10.1037/a0030915

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, *32*(3), 391–421. https://doi.org/10.1080/13854046.2017.1406146

Lovett, B. J., & Harrison, A. G. (2021). Assessing adult ADHD: New research and perspectives. *Journal of Clinical and Experimental Neuropsychology*, *43*(4), 333–339. https://doi.org/10.1080/13803395.2021.1950640

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*(7), 1204–1237. https://doi.org/10.1080/13854046.2010.514290

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2012). Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of clinical Neuropsychology*, *27*(1), 10–22. https://doi.org/10.1093/arclin/acr084

Ord, A. S., Miskey, H. M., Lad, S., Richter, B., Nagy, K., & Shura, R. D. (2021). Examining embedded validity indicators in Conners continuous performance test-3 (CPT-3). *The Clinical Neuropsychologist*, *35*(8), 1426–1441. https://doi.org/10.1080/13854046.2020.1751301

Ord, J. S., Boettcher, A. C., Greve, K. W., & Bianchini, K. J. (2010). Detection of malingering in mild traumatic brain injury with the Conners' continuous performance test–II. *Journal of Clinical and Experimental Neuropsychology*, *32*(4), 380–387. https://doi.org/10.1080/13803390903066881

Schroeder, R. W., Martin, P. K., Heinrichs, R. J., & Baade, L. E. (2018). Research methods in performance validity testing studies: Criterion grouping approach impacts study outcomes. *The Clinical Neuropsychologist*, *33*(3), 466–477. https://doi.org/10.1080/13854046.2018.1484517

Scimeca, L. M., Holbrook, L., Rhoads, T., Cerny, B. M., Jennette, K. J., Resch, Z. J., Obolsky, M. A., Ovsiew, G. P., & Soble, J. R. (2021). Examining Conners continuous performance test-3 (CPT-3) embedded performance validity indicators in an adult clinical sample referred for ADHD evaluation. *Developmental Neuropsychology*, *46*(5), 347–359. https://doi.org/10.1080/87565641.2021.1951270

Sharland, M. J., Waring, S. C., Johnson, B. P., Taran, A. M., Rusin, T. A., Pattock, A. M., & Palcher, J. A. (2018). Further examination of embedded performance validity indicators for the Conners' continuous performance test and brief test of attention in a large outpatient clinical sample. *The Clinical Neuropsychologist*, *32*(1), 98–108. https://doi.org/10.1080/13854046.2017.1332240

Sherman, E., 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

Shura, R. D., Miskey, H. M., Rowland, J. A., Yoash-Gantz, R. E., & Denning, J. H. (2016). Embedded performance validity measures with post deployment veterans: Cross-validation and efficiency with multiple measures. *Applied Neuropsychology: Adult*, *23*(2), 94–104. https://doi.org/10.1080/23279095.2015.1014556

Suhr, J. A., & Berry, D. T. R. (2017). The importance of assessing for validity of symptom report and performance in attention deficit/hyperactivity disorder (ADHD): Introduction to the special section on noncredible presentation in ADHD. *Psychological Assessment*, *29*(12), 1427–1428. https://doi.org/10.1037/pas0000535

Sweet, J. J., Benson, L. M., Nelson, N. W., & Moberg, P. J. (2015). The American Academy of Clinical Neuropsychology, National Academy of Neuropsychology, and Society for Clinical Neuropsychology (APA division 40) 2015 TCN professional practice and 'salary survey': Professional practices, beliefs, and incomes of U.S. neuropsychologists. *The Clinical Neuropsychologist*, *29*(8), 1069–1162. https://doi.org/10.1080/13854046.2016.1140228

Sweet, J. J., Klipfel, K. M., Nelson, N. W., & Moberg, P. J. (2021). Professional practices, beliefs, and incomes of U.S. neuropsychologists: The AACN, NAN, SCN 2020 practice and "salary survey". *The Clinical Neuropsychologist*, *35*(1), 7–80. https://doi.org/10.1080/13854046.2020.1849803

Tombaugh, T. N. (1996). *TOMM: Test of memory malingering*. Multi-Health Systems.

White, D. J., Korinek, D., Bernstein, M. T., Ovsiew, G. P., Resch, Z. J., & Soble, J. R. (2020). 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

Winter, D., & Braw, Y. (2022). Validating embedded validity indicators of feigned ADHD-associated cognitive impairment using the MOXO-d-CPT. *Journal of Attention Disorders*. Advance online publication. https://doi.org/10.1177/10870547221112947

Author Biographies

Anthony Robinson is a clinical psychology doctoral student at Louisiana State University in Baton Rouge, LA.

Christopher Reed is a clinical psychology doctoral student at Louisiana State University in Baton Rouge, LA.

Katrail Davis is a clinical psychology doctoral student at Louisiana State University in Baton Rouge, LA.

Ross Divers is a clinical psychology doctoral student at Louisiana State University in Baton Rouge, LA.

Luke Miller is a clinical psychology doctoral student at Louisiana State University in Baton Rouge, LA.

Laszlo A. Erdodi is a professor of psychology at the University of Windsor in Ontario, Canada.

Matthew Calamia is an associate professor of psychology and director of the Psychological Services Center at Louisiana State University in Baton Rouge, LA.