Psychological Assessment

ISSN: 1040-3590

2022, Vol. 34, No. 7, 697–703
https://doi.org/10.1037/pas0001137

BRIEF REPORT

A Comparison of Embedded Validity Indicators From the Stroop Color and Word Test Among Adults Referred for Clinical Evaluation of Suspected or Confirmed Attention-Deficit/Hyperactivity Disorder

Humza Khan[1, 2], Andrew A. Rauch[1, 3], Maximillian A. Obolsky[1, 4], Haley Skymba[1, 5], Kearston C. Barwegen[1, 6], Amanda M. Wisinger[1, 7], Gabriel P. Ovsiew[1], Kyle J. Jennette[1], Jason R. Soble[1, 8], and Zachary J. Resch[1]

[1] Department of Psychiatry, University of Illinois College of Medicine
[2] Department of Psychology, Illinois Institute of Technology
[3] Department of Psychology, Loyola University
[4] Department of Psychology, Roosevelt University
[5] Department of Psychology, University of Illinois at Urbana-Champaign
[6] Department of Psychology, Wheaton College
[7] Department of Psychology, Chicago School of Professional Psychology
[8] Department of Neurology, University of Illinois College of Medicine

This study investigated the utility of four Stroop Color and Word Test (SCWT) indices, including the raw score and T score for the word reading (WR) and color naming (CN) trials, as embedded performance validity tests (PVTs) within a sample referred for evaluation of suspected or known attention-deficit/hyperactivity disorder (ADHD). Data were analyzed from a final sample of 317 patients consecutively referred for ADHD evaluation, which was divided into groups with invalid (n = 43; 14%) and valid neuropsychological test performance (n = 274; 86%). A subset of the valid group with confirmed ADHD diagnoses (n = 226; 71%) were also analyzed separately. Classification accuracy for the overall valid sample was in the acceptable range (AUCs = .757–.794), with optimal cut scores of WR raw ≤75 (54% sensitivity/90% specificity), WR T score ≤28 (54% sensitivity/88% specificity), CN raw ≤57 (42% sensitivity/90% specificity), and CN T score ≤30 (40% sensitivity/90% specificity). Classification accuracy was also in the acceptable range for the ADHD-confirmed subgroup (AUCs = .750–.790), with optimal cut scores of WR Raw ≤75 (54% sensitivity/89% specificity), WR T score ≤28 (54% sensitivity/87% specificity), CN Raw ≤57 (42% sensitivity/90% specificity), and CN T score ≤30 (40% sensitivity/90% specificity). These findings indicate that embedded PVTs derived from the SCWT, particularly those derived from the WR trial, are effective measures for determining validity status in samples with suspected or confirmed ADHD.

Public Significance Statement

This study cross-validated four indices within the Stroop Color and Word Test (i.e., the word reading raw score and T score and the color naming raw score and T score) as embedded performance validity tests in a sample referred for ADHD evaluation. Findings suggest that these indices function as embedded validity indicators in samples with suspected or confirmed ADHD.

Keywords: Stroop Color and Word Test, performance validity, psychometrics, attention/deficit hyperactivity disorder, ADHD

This article was published Online First March 31, 2022.
Zachary J. Resch https://orcid.org/0000-0002-3199-6168
Humza Khan played lead role in formal analysis and equal role in investigation, methodology and writing of original draft. Andrew A. Rauch played equal role in writing of original draft. Maximillian A. Obolsky played equal role in writing of original draft. Haley Skymba played equal role in writing of original draft. Kearston C. Barwegen played equal role in writing of original draft. Amanda M. Wisinger played equal role in writing of original draft. Gabriel P. Ovsiew played supporting role in writing of original draft and equal role in data curation. Kyle J. Jennette played equal role in writing of original draft. Jason R. Soble played supporting role in conceptualization and methodology and equal role in data curation, investigation and writing of original draft. Zachary J. Resch played lead role in methodology, project administration and supervision and equal role in formal analysis, investigation and writing of original draft.

This study was not preregistered. Materials and analysis code for this study are available by emailing the corresponding author.

Correspondence concerning this article should be addressed to Zachary J. Resch, Department of Psychiatry, University of Illinois College of Medicine, 912 S. Wood Street (MC 913), Chicago, IL 60612, United States. Email: zresch2@uic.edu

Neuropsychologists frequently receive referrals to assist with diagnostic clarification and treatment planning of adults with suspected attention-deficit/hyperactivity disorder (ADHD). High rates of invalid neuropsychological test performance have been found among adults presenting for ADHD evaluation based on both empirical studies (i.e., 9%–50%; Leib et al., 2021; Marshall et al., 2016; Suhr et al., 2008; Sullivan et al., 2007; White et al., 2022) as well as surveys of practicing clinical neuropsychologists (i.e., 20%; Martin & Schroeder, 2020). This is somewhat unsurprising given the multiple potential external incentives associated with an ADHD diagnosis, including access to psychostimulant medications, academic and occupational accommodations, and standardized testing accommodations (Suhr & Berry, 2017). As such, objective assessment of performance validity is a critical component of adult ADHD evaluations. This mirrors broader practice standards calling for the routine administration of multiple freestanding and embedded performance validity tests (PVTs) during all neuropsychological evaluations to ensure test results are a valid reflection of an examinee's true cognitive abilities and to minimize errors associated with inaccurate diagnosis (Sweet et al., 2021).

To meet current standards of practice regarding validity assessment, the past 2 decades have produced a proliferation of research cross-validating freestanding and embedded PVTs in various clinical populations (see Soble et al., 2021 for a review). Many of these available PVTs are or appear to be memory-based, whereas far fewer nonmemory-based PVTs have been developed. Among embedded PVTs, a few recent studies have investigated various indices from the Stroop Color and Word Test (SCWT; Golden, 1978) and found them to be effective in accurately detecting invalid test performance.[1] For instance, Shura et al. (2016) reported that the SCWT word reading (WR) raw score had .23 sensitivity/.90 specificity among postdeployment veterans. Lee and colleagues reported acceptable sensitivities (i.e., .40–.49) with ≥.90 specificity across the WR, color naming (CN), and color–word interference (CW) trials among a mixed clinical and forensic sample. More recently, White et al. (2020) investigated these three SCWT indices within a mixed clinical sample with and without cognitive impairment and reported .31–.54 sensitivity/≥.89 specificity, with the WR T score generally having the more robust and stable sensitivity/specificity values across the cognitively intact and impaired groups.

As the SCWT assesses both processing speed and executive functioning, and various formats of the task are commonly included in neuropsychological testing batteries for ADHD (Lee et al., 2019; Rabin et al., 2016), embedded PVTs derived from this test paradigm may be especially valuable in allowing for continuous assessment of validity status (Boone, 2009) without necessitating the additional costs, time, and patient burden associated with adding standalone PVTs to a test battery. Moreover, as ADHD is not traditionally thought of as a disorder of memory by the lay public, these SCWT indices may offer unique utility as nonmemory-based PVTs. However, cross-validation of these SCWT embedded PVTs among adult ADHD-specific populations is critical to ensure (a) that these indices are able to accurately differentiate valid from invalid performance and (b) that they identify the optimal cut score(s) for these indices that maximize sensitivity while maintaining adequate specificity in this clinical population. Given that specificity is prioritized over sensitivity with PVTs in order to avoid inflated false positive classifications, embedded measures often have low–to-moderate sensitivity when maintaining adequate specificity levels (i.e., ≥.90), with sensitivity values of ≥.40 generally considered acceptable for embedded validity indicators (Boone, 2013). Accordingly, this study assessed the effectiveness of the SCWT for detecting performance invalidity among a large sample of adults clinically referred for ADHD evaluation. We hypothesized that following previous research on using the SCWT as an embedded PVT (e.g., Shura et al., 2016; White et al., 2020), all SCWT indices will accurately differentiate valid from invalid performance, and that WR scores, in particular, would evidence the most robust sensitivity for detecting performance invalidity.

Method

Participants

Data for this cross-sectional study were collected from 323 adults consecutively referred for outpatient neuropsychological evaluation of known or suspected ADHD from 2018 to 2021 at a Midwestern academic medical center and consented to including their test data as part of an ongoing Institutional Review Board-approved study. This study was not preregistered. Examinees were referred for evaluation by their medical providers (most commonly their psychiatrist or primary care/family medicine physician) specifically for the purposes of diagnostic clarification and/or updated treatment planning related to suspected ADHD. Notably, all examinees underwent a uniform, multimethod diagnostic protocol to evaluate for the presence/absence of ADHD. This protocol included: (a) medical/psychiatric record review (including review of previous evaluation/diagnostic workup of ADHD, when available); (b) a semistructured clinical interview to gather all relevant background information (e.g., ADHD symptom onset and course; associated functional impairment; medical, psychiatric, developmental, and psychosocial history) and thoroughly assess formal diagnostic criteria for ADHD and possible psychiatric comorbidities; (c) administration of the clinical assessment of attention deficit-adult (CAT-A; Bracken & Boatwright, 2005), which contains symptom validity scales to identify noncredible ADHD symptom reporting and provides normative-based qualification of childhood and current ADHD symptomatology; (d) administration of a standardized neuropsychological test battery to comprehensively assess examinees' cognition across all major domains; and (e) administration of a validity-controlled personality inventory [i.e., Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008)] to objectively assess for active comorbid psychopathology. All evaluations were conducted in person, and all ADHD diagnoses were rendered by a board-certified clinical neuropsychologist. All patients reported English as their primary language, and test administration for each patient was completed in English.

Among the testing battery, all patients were administered seven freestanding and embedded PVTs used as a reference standard in the present study, including the Dot Counting Test (DCT; Boone et al., 2002); Trail Making Test-part A (TMT-A; White et al., 2020); letter fluency (e.g., F/A/S; White et al., 2020); Reliable Digit Span

[1] Of note, embedded validity indicators have also been derived from and validated using other Stroop paradigms, including the Comalli version (Arentsen et al., 2013) and Delis–Kaplan Executive Function System Color–Word Interference (Eglit et al., 2020; Erdodi et al., 2018), but these will not be reviewed here given the focus of our article is the traditional SCWT.

(RDS; Schroeder et al., 2012); Rey Auditory Verbal Learning Test (RAVLT; Pliskin et al., 2021); Brief Visuospatial Memory Test-Revised (BVMT-R; Bailey et al., 2018; Resch et al., 2020); and the Rey 15-Item Test (RFIT; Poynter et al., 2019). Six patients were not administered one of the criterion PVTs (i.e., $n = 5$ missing the RFIT and $n = 1$ missing the DCT), and these cases were therefore excluded from subsequent analysis. Patients with one or fewer PVT failures constituted the valid group ($n = 274$), whereas those with two or more PVT failures constituted the invalid group ($n = 43$). This grouping method is consistent with current practice standards as well as the revised criteria for identifying invalid neuropsychological test profiles (Sherman et al., 2020; Sweet et al., 2021), and is further empirically supported by recent findings suggesting that the inclusion of individuals with a single PVT failure as performing validly is an acceptable PVT research methodology (Jennette et al., 2021; Rhoads et al., 2021). Of the 274 in the valid group, 226 cases met formal *Diagnostic and Statistical Manual of Mental Disorders-5th Edition* (DSM-5; American Psychiatric Association, 2013) criteria for ADHD and were therefore assigned to the valid-confirmed ADHD group.

Table 1 includes sample diagnostic composition presented by subgroup (i.e., valid-ADHD referrals, valid-confirmed ADHD, and invalid). The valid sample was 61% female ($n = 198$) and 39% male ($n = 125$) and varied in terms of age ($M = 27.70$ years, $SD = 6.67$; range = 18–60) and education ($M = 15.71$ years, $SD = 2.05$; range = 8–20). The ethnoracial breakdown of the sample was 46% non-Hispanic White ($n = 149$), 24% non-Hispanic Black ($n = 77$), 15% Hispanic ($n = 47$), 10% Asian/Pacific Islander ($n = 33$), and 5% other race/ethnicity ($n = 17$). Regarding demographic differences across groups (Table 2), comparison of overall valid (all ADHD referrals) and invalid group revealed that the valid group had approximately 1 more year of education, $F(1, 316) = 6.83$, $p = .009$, $\eta_p^2 = .021$, and a greater proportion of women, $\chi^2(1, N = 317) = 4.52$, $p = .03$, than the invalid group. Differences in age and ethnoracial composition differences between the overall valid and invalid groups were not statistically significant (ps > .05). Among the subset of the valid group with confirmed ADHD, a similar pattern emerged whereby the valid ADHD-confirmed subgroup had roughly one more year of education, on average, $F(1, 268) = 7.32$, $p = .007$, $\eta_p^2 = .027$, and a greater proportion of women, $\chi^2(1, N = 269) = 4.25$, $p = .04$, than the invalid group. The valid-confirmed ADHD subgroup also had a significantly different ethnoracial composition than the invalid group, $\chi^2(4, N = 269) = 10.82$, $p = .03$. No groups differed significantly in terms of age (ps > .05). Finally, although all testing was completed in English, approximately 31% of the sample identified as bilingual. Monolingual English speakers and bilingual English/Spanish speakers performed equivalently ($p = .17$–$.95$) across all four SCWT indices examined (see Measures section below), with negligible effect sizes ($\eta_p^2 = .000$–$.006$). Therefore, no language-based performance differences were found for the SCWT.

Measures

Stroop Color and Word Test (Golden, 1978)

The SCWT consists of three trials. Within the first trial (WR), an individual must read as many color names as possible. Within the second trial (CN), an individual must name the color of the ink in which the letter Xs are printed. Within the third, CW interference trial, an individual must name the color of the ink in which different color names are written (e.g., "green" written in red ink) rather than the word itself. Each trial has a 45-s time limit, during which the examinee must try to read as many words or name as many colors as possible. This study examined the raw and demographically corrected T scores for the WR and CN trials.

Data Analytic Plan

Descriptive statistics were computed for each SCWT embedded PVT, and chi-square tests and analyses of variance were conducted to assess performance differences between the valid-ADHD referral/invalid groups and valid-confirmed ADHD/invalid groups.

Table 1
Demographic Characteristics for the Sample Stratified by Group

Demographic factors	Valid-all ADHD referrals ($n = 274$)		Valid-subgroup with confirmed ADHD ($n = 226$)		Invalid ($n = 43$)	
	M (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range
Age	27.92 (6.75)	18–60	27.98 (6.81)	18–60	26.26 (6.02)	18–43
Education	15.82 (2.02)	9–20	15.87 (2.03)	9–20	14.95 (2.12)	8–18
	N	%	*N*	%	*N*	%
Sex						
Male	100	26	83	37	23	53
Female	174	64	143	63	20	47
Race/ethnicity						
White	131	48	113	50	16	37
Black	62	23	49	22	13	30
Hispanic	35	13	26	12	11	26
Asian/Pacific Islander	29	11	3	10	3	7
Other	17	6	15	6	—	—

Note. ADHD = attention-deficit/hyperactivity disorder.

Table 2
Spearman Correlations Between the SCWT Embedded Validity Indicators and the Criterion PVTs Among Validly Performing Patients

Performance validity tests	SCWT WR raw	SCWT WR T score	SCWT CN raw	SCWT CN T score
SCWT WR raw	—	.965**	.631**	.599**
SCWT WR T score	.965**	—	.633**	.641**
SCWT CN raw	.631**	.633**	—	.982**
SCWT CN T score	.599**	.641**	.982**	—
Trail making test-part A	.227**	.236**	.204**	.214**
Letter fluency (F/A/S T score)	.196**	.207**	.188**	.199**
Reliable digit span	.220**	.173**	.169**	.142*
RAVLT effort score	.040	.027	.100	.094
BVMT-R RD	−.047	−.040	.093	.089
Dot counting test	−.365**	−.331**	−.402**	−.373**
Rey 15-item test recall + recognition	.026	−.010	.019	−.014

Note. $n = 274$. SCWT = Stroop Color and Word Test; WR = word reading trial; CN = color naming trial; RAVLT = Rey Auditory Verbal Learning Test; BVMT-R RD = Brief Visuospatial Memory Test-Revised recognition discrimination.
* $p < .05$. ** $p < .01$.

Spearman correlations were conducted to examine relationships among the different SCWT embedded PVT indices as well as their associations with the criterion PVTs. Finally, receiver operating characteristic (ROC) curve analyses were performed to examine the overall classification accuracy of each PVT index and to determine optimal cut scores that maximize sensitivity and specificity. The ROC curve analyses were first conducted for the overall sample (i.e., valid-ADHD referral/invalid) as well as for the subset of the sample with confirmed ADHD diagnoses (i.e., valid-confirmed ADHD/invalid) for the purposes of relative comparison and to determine if the accuracies of the SCWT embedded PVTs differ among the more homogenous group with diagnosed ADHD. For the ROC curve analyses, classification accuracy via areas under the curve (AUCs) was interpreted as poor (.50–.69), acceptable (.70–.79), excellent (.80–.89), or outstanding (≥.90; Hosmer et al., 2013). Materials and analysis code for this study are available by emailing the corresponding author.

Results

Correlations between the SCWT embedded validity indicators are presented in the upper portion of Table 2. All SCWT embedded indices were intercorrelated, with medium to large effects. The correlations were robust between most indices and with CN raw score and T score being more strongly correlated than WR raw score and T score. Regarding the correlations between SCWT embedded validity indicators and criterion PVTs (see the lower portion of Table 2), associations ranged from nonsignificant to medium effects. All SCWT indices were negatively correlated with the DCT, whereas there were positive correlations with TMT-A, F/A/S, and RDS and nonsignificant correlations with RAVLT, BVMT-R, and RFIT.

When comparing the overall valid group (i.e., all ADHD referrals) to the invalid group (Table 3), classification accuracies for the SCWT embedded validity indicators were statistically significant and within the acceptable range (AUCs = .757–.794). WR raw score emerged as the validity index with the best combination of sensitivity and specificity in these analyses, demonstrating .53 sensitivity/.90 specificity at an optimal cut scores of ≤75. For the ROC analyses between the valid-confirmed ADHD subgroup and the invalid group (Table 4), classification accuracies were statistically significant for all indices and again within the acceptable range (AUCs = .750–.790). WR raw score produced the best psychometrics among this subset, with .54 sensitivity/.89 specificity at its optimal cut score of ≤75. Compared to the results above, CN T score showed the greatest changes in cut score (≤32 vs. ≤30) and sensitivity (.47 vs. .40), with all other indices remaining relatively consistent.

Discussion

This study served as cross-validation of SCWT embedded validity indicators in an adult clinical sample referred for evaluation of suspected ADHD. Overall, all four SCWT indices showed acceptable classification accuracy among individuals referred for ADHD, as well as the subset with confirmed ADHD diagnoses. In both sets of analyses, the SCWT WR raw score yielded the most robust combination of sensitivity and specificity at an optimal cut score of ≤75 (.53–.54 sensitivity/.89–.90 specificity). Taken together, these results demonstrated that individual SCWT indices, especially those derived from the WR trial, have utility as embedded PVTs for use in clinical ADHD evaluations.

The present results generally corroborated previous findings of SCWT indices showing promise as embedded validity indicators. Namely, sensitivities and specificities obtained in the present study sample (WR: .54/.88; CN: .40–.42/.90) are remarkably comparable to those yielded in a mixed neuropsychiatric sample (WR: .54/.90–.91; CN: .39–.46/.90; White et al., 2020). The relative psychometric stability across such distinct samples bodes well for using SCWT indices as embedded PVTs in different clinical and research settings. The key difference between these study findings was that more liberal cut scores were recommended among the patients referred for ADHD evaluation, which is not surprising given that more conservative cut scores are often necessitated to offset higher rates of false positive results in samples with more cognitive impairment. Moreover, these samples differed demographically, with the current sample consisting of younger, more highly educated individuals than those evaluated by White et al. (2020). This latter observation may explain why the WR T score was the preferred validity metric from the SCWT in the prior study in which there was a wider range of ages and educational attainment, whereas the present results suggested that the WR raw score yielded slightly better specificity than the T score. The WR and CN trials likely work

Table 3

Receiver Operating Characteristic Curve Analysis of the SCWT Indices Among Those Referred for ADHD Evaluation

Validity index	AUC	Cut score	SN	SP	10% base rate PPV	10% base rate NPV	20% base rate PPV	20% base rate NPV	30% base rate PPV	30% base rate NPV	40% base rate PPV	40% base rate NPV
SCWT WR raw	.779***	≤74	.465	.901	0.34	0.94	0.54	0.87	0.67	0.80	0.76	0.72
		≤75	**.535**	**.901**	**0.38**	**0.95**	**0.57**	**0.89**	**0.70**	**0.82**	**0.78**	**0.74**
		≤77	.535	.891	0.35	0.95	0.55	0.88	0.68	0.82	0.77	0.74
		≤78	.558	.876	0.33	0.95	0.53	0.89	0.66	0.82	0.75	0.75
SCWT WR T score	.757***	≤27	.442	.883	0.30	0.93	0.49	0.86	0.62	0.79	0.72	0.70
		≤28	**.535**	**.876**	**0.32**	**0.94**	**0.52**	**0.88**	**0.65**	**0.81**	**0.74**	**0.74**
		≤29	.535	.869	0.31	0.94	0.51	0.88	0.64	0.81	0.73	0.74
		≤30	.581	.832	0.28	0.95	0.46	0.89	0.60	0.82	0.70	0.75
SCWT CN raw	.794***	≤56	.395	.905	0.32	0.93	0.51	0.86	0.64	0.78	0.73	0.69
		≤57	**.419**	**.901**	**0.32**	**0.93**	**0.51**	**0.86**	**0.64**	**0.78**	**0.74**	**0.70**
		≤58	.442	.887	0.30	0.93	0.49	0.86	0.63	0.79	0.72	0.70
		≤59	.465	.869	0.28	0.94	0.47	0.87	0.60	0.79	0.70	0.71
SCWT CN T score	.766***	≤29	.372	.909	0.31	0.93	0.51	0.85	0.64	0.77	0.73	0.68
		≤30	**.395**	**.898**	**0.30**	**0.93**	**0.49**	**0.86**	**0.62**	**0.78**	**0.72**	**0.69**
		≤31	.419	.883	0.28	0.93	0.47	0.86	0.61	0.78	0.70	0.70
		≤32	.465	.861	0.27	0.94	0.46	0.87	0.59	0.79	0.69	0.71

Note. Valid-ADHD referral group (n = 274), invalid group (n = 43). Cut-scores with maximum diagnostic accuracy (i.e., maximizing sensitivity while maintaining acceptable specificity) have been bolded. SCWT = Stroop Color and Word Test; ADHD = attention-deficit/hyperactivity disorder; AUC = area under the curve; SN = sensitivity; SP = specificity; PPV = positive predictive value; NPV = negative predictive value; WR = word reading trial. CN = color naming trial.
$^{***}p < .001$.

well as embedded PVTs largely due to fact that the underlying cognitive abilities are overlearned and likely more robust to cognitive impairment. A notable caveat is that these measures have not yet been cross-validated in patient populations or specific diagnostic groups with significant processing speed impairment, which may adversely affect their ability to assess performance validity.

Several study limitations were noteworthy. First and foremost, informant report data were generally not available to corroborate adult ADHD diagnoses. Although this limitation is frequently characteristic of adult ADHD assessments, the absence of informant data may have resulted in increased false positive ADHD diagnoses. Next, the sample tended to have higher than average educational attainment (i.e., 15.66 years) as most examinees were college undergraduates, graduate students, or professional students, which may limit generalizability to samples with less education. This particular limitation may be more germane to the cut scores based

Table 4

Receiver Operating Characteristic Curve Analysis of the SCWT Indices Comparing Those With Confirmed ADHD Diagnoses to Invalid Performers

Validity index	AUC	Cut score	SN	SP	10% base rate PPV	10% base rate NPV	20% base rate PPV	20% base rate NPV	30% base rate PPV	30% base rate NPV	40% base rate PPV	40% base rate NPV
SCWT WR raw	.775***	≤74	.465	.894	0.33	0.94	0.52	0.87	0.65	0.80	0.75	0.71
		≤75	**.535**	**.894**	**0.36**	**0.95**	**0.56**	**0.88**	**0.68**	**0.82**	**0.77**	**0.74**
		≤77	.535	.881	0.33	0.94	0.53	0.88	0.66	0.82	0.75	0.74
		≤78	.558	.863	0.31	0.95	0.50	0.89	0.64	0.82	0.73	0.75
SCWT WR T score	.750***	≤27	.442	.876	0.28	0.93	0.47	0.86	0.60	0.79	0.70	0.70
		≤28	**.535**	**.867**	**0.31**	**0.94**	**0.50**	**0.88**	**0.63**	**0.81**	**0.73**	**0.74**
		≤29	.535	.863	0.30	0.94	0.49	0.88	0.63	0.81	0.72	0.74
		≤30	.581	.823	0.27	0.95	0.45	0.89	0.58	0.82	0.69	0.75
SCWT CN raw	.790***	≤56	.395	.903	0.31	0.93	0.50	0.86	0.64	0.78	0.73	0.69
		≤57	**.419**	**.898**	**0.31**	**0.93**	**0.51**	**0.86**	**0.64**	**0.78**	**0.73**	**0.70**
		≤58	.442	.881	0.29	0.93	0.48	0.86	0.61	0.79	0.71	0.70
		≤59	.465	.863	0.27	0.94	0.46	0.87	0.59	0.79	0.69	0.71
SCWT CN T score	.761***	≤29	.372	.903	0.30	0.93	0.49	0.85	0.62	0.77	0.72	0.68
		≤30	**.395**	**.898**	**0.30**	**0.93**	**0.49**	**0.86**	**0.62**	**0.78**	**0.72**	**0.69**
		≤31	.419	.881	0.28	0.93	0.47	0.86	0.60	0.78	0.70	0.69
		≤32	.465	.858	0.27	0.94	0.45	0.87	0.58	0.79	0.69	0.71

Note. Valid-confirmed ADHD group (n = 226), invalid group (n = 43). Cut-scores with maximum diagnostic accuracy (i.e., maximizing sensitivity while maintaining acceptable specificity) have been bolded. SCWT = Stroop Color and Word Test; ADHD = attention-deficit/hyperactivity disorder; AUC = area under the curve; SN = sensitivity; SP = specificity; PPV = positive predictive value; NPV = negative predictive value; WR = word reading trial. CN = color naming trial.
$^{***}p < .001$.

on the raw score versus the T score, which accounts for educational attainment. Third, despite the use of seven criterion PVTs, most were embedded validity indicators, which are typically considered less psychometrically robust than freestanding PVTs (e.g., Ovsiew et al., 2020; Pliskin et al., 2021) and may affect the rates of false positive and false negative errors when determining validity classification. Despite the utility of embedded PVTs, inclusion of freestanding PVTs remains warranted given their generally higher sensitivity (Soble et al., 2021).

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders-Fifth edition (DSM-5)*. American Psychiatric Publishing.

Arentsen, T. J., Boone, K. B., Lo, T. T., Goldberg, H. E., Cottingham, M. E., Victor, T. L., Ziegler, E., & Zeller, M. A. (2013). Effectiveness of the Comalli Stroop Test as a measure of negative response bias. *The Clinical Neuropsychologist*, *27*(6), 1060–1076. https://doi.org/10.1080/13854046.2013.803603

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*, 895–900. https://doi.org/10.1093/arclin/acx111

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

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. (2013). *Clinical practice of forensic neuropsychology: An evidence-based approach*. Guilford Press.

Boone, K. B., Lu, P., & Herzberg, D. S. (2002). *The dot counting test manual*. Western Psychological Services.

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

Eglit, G. M. L., Jurick, S. M., Delis, D. C., Filoteo, J. V., Bondi, M. W., & Jak, A. J. (2020). Utility of the D-KEFS Color Word Interference Test as an embedded measure of performance validity. *The Clinical Neuropsychologist*, *34*(2), 332–352. https://doi.org/10.1080/13854046.2019.1643923

Erdodi, L. A., Sagar, S., Seke, K., Zuccato, B. G., Schwartz, E. S., & Roth, R. M. (2018). The Stroop test as a measure of performance validity in adults clinically referred for neuropsychological assessment. *Psychological Assessment*, *30*(6), 755–766. https://doi.org/10.1037/pas0000525

Golden, C. J. (1978). *The Stroop Color and Word Test: A manual for clinical and experimental uses*. Stoelting.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley and Sons. https://doi.org/10.1002/9781118548387

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. A., & 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*, *16*(1), 1–19. https://doi.org/10.1080/13854046.2021.1900398

Lee, C., Landre, N., & Sweet, J. J. (2019). Performance validity on the Stroop Color and Word Test in a mixed forensic and patient sample. *The Clinical Neuropsychologist*, *33*(8), 1403–1419. https://doi.org/10.1080/13854046.2019.1594385

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*, *20*(2), 1–8. https://doi.org/10.1080/23279095.2021.1894150

Marshall, P. S., Hoelzle, J. B., Heyerdahl, D., & Nelson, N. W. (2016). The impact of failing to identify suspect effort in patients undergoing adult attention-deficit/hyperactivity disorder (ADHD) assessment. *Psychological Assessment*, *28*(10), 1290–1302. https://doi.org/10.1037/pas0000247

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*, 717–725. https://doi.org/10.1093/arclin/acaa017

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

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 8 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

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 cut-offs for Rey 15-item recall and recognition. *Archives of Clinical Neuropsychology*, *34*(8), 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*, *29*(2), 252–261. https://doi.org/10.1080/23279095.2020.1742718

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

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. https://doi.org/10.1177/1073191111428764

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

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

Soble, J. R., Webber, T. A., & Bailey, K. C. (2021). An overview of common stand-alone and embedded PVTs for the practicing clinician: Cutoffs, classification accuracy, and administration times. In R. W. Schroeder &

P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 126–149). Guilford.

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. https://doi.org/10.1016/j.acn.2008.05.003

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

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. https://doi.org/10.1080/09084280701509083

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., & the Conference Participants. (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 Neuropsychologist*, *35*(6), 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

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

White, D. J., Ovsiew, G. P., Rhoads, T., Resch, Z. J., Lee, M., Oh, A. J., & Soble, J. R. (2022). The divergent roles of symptom and performance validity in the assessment of ADHD. *Journal of Attention Disorders*, *26*(1), 101–108. https://doi.org/10.1177/1087054720964575

Received November 9, 2021
Revision received January 24, 2022
Accepted March 1, 2022 ■