Routledge
Taylor & Francis Group

Journal of Clinical and Experimental Neuropsychology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ncen20

Cross-validating the Clinical Assessment of Attention Deficit–Adult symptom validity scales for assessment of attention deficit/hyperactivity disorder in adults

John-Christopher A. Finley, Brian M. Cerny, Julia M. Brooks, Maximillian A. Obolsky, Aya Haneda, Gabriel P. Ovsiew, Devin M. Ulrich, Zachary J. Resch & Jason R. Soble

To cite this article: John-Christopher A. Finley, Brian M. Cerny, Julia M. Brooks, Maximillian A. Obolsky, Aya Haneda, Gabriel P. Ovsiew, Devin M. Ulrich, Zachary J. Resch & Jason R. Soble (2024) Cross-validating the Clinical Assessment of Attention Deficit–Adult symptom validity scales for assessment of attention deficit/hyperactivity disorder in adults, Journal of Clinical and Experimental Neuropsychology, 46:2, 111-123, DOI: 10.1080/13803395.2023.2283940

To link to this article: https://doi.org/10.1080/13803395.2023.2283940

JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY
2024, VOL. 46, NO. 2, 111–123
https://doi.org/10.1080/13803395.2023.2283940

Routledge
Taylor & Francis Group

Cross-validating the Clinical Assessment of Attention Deficit–Adult symptom validity scales for assessment of attention deficit/hyperactivity disorder in adults

John-Christopher A. Finley[a], Brian M. Cerny[b,c], Julia M. Brooks[c,d], Maximillian A. Obolsky[c,e], Aya Haneda[e], Gabriel P. Ovsiew[c], Devin M. Ulrich[c], Zachary J. Resch[c] and Jason R. Soble[c,f]

[a]Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School, Chicago, IL, USA; [b]Department of Psychology, Illinois Institute of Technology Chicago, Chicago, IL, USA; [c]Department of Psychiatry, University of Illinois College of Medicine, Chicago, IL, USA; [d]Department of Psychology, University of Illinois College of Medicine, Chicago, IL, USA; [e]Department of Psychology, Roosevelt University, Chicago, IL, USA; [f]Department of Neurology, University of Illinois College of Medicine, Chicago, IL, USA

ABSTRACT

Introduction: The Clinical Assessment of Attention Deficit-Adult is among the few questionnaires that offer validity indicators (i.e., Negative Impression [NI], Infrequency [IF], and Positive Impression [PI]) for classifying underreporting and overreporting of attention-deficit/hyperactivity disorder (ADHD) symptoms. This is the first study to cross-validate the NI, IF, and PI scales in a sample of adults with suspected or known ADHD.

Method: Univariate and multivariate analyses were conducted to examine the independent and combined value of the NI, IF, and PI scores in predicting invalid symptom reporting and neurocognitive performance in a sample of 543 adults undergoing ADHD evaluation.

Results: The NI scale demonstrated better classification accuracy than the IF scale in discriminating patients with and without valid scores on measures of overreporting. Only NI scores significantly predicted validity status when used in combination with IF scores. Optimal cut-scores for the NI (≤51; 30% sensitivity / 90% specificity) and IF (≥4; 18% sensitivity / 90% specificity) scales were consistent with those reported in the original manual; however, these indicators poorly discriminated patients with invalid and valid neurocognitive performance. The PI scale demonstrated acceptable classification accuracy in discriminating patients with invalid and valid scores on measures of underreporting, albeit with an optimal cut-score (≥27; 36% sensitivity / 90% specificity) lower than that described in the manual.

Conclusion: Findings provide preliminary evidence of construct validity for these scales as embedded validity indicators of symptom overreporting and underreporting. However, these scales should not be used to guide clinical judgment regarding the validity of neurocognitive test performance.

ARTICLE HISTORY
Received 12 June 2023
Accepted 9 November 2023

KEYWORDS
CAT-A; symptom validity; ADHD; overreporting; underreporting

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder primarily diagnosed in childhood with symptoms that may persist into adulthood (Fayyad et al., 2017). Assessment of ADHD symptoms is now one of the most common reasons for neuropsychological evaluation in adult outpatient settings (Rabin et al., 2016; Schroeder et al., 2019). Neuropsychologists typically assess symptoms using clinical interview and self-report questionnaires in addition to neurocognitive testing. However, fabrication or exaggeration of ADHD symptoms is common in adult outpatient settings for reasons related to potential secondary gain, including access to stimulant medication and academic and standardized testing accommodations (Sagar et al., 2017). Up to 22%-48% of patients undergoing ADHD assessment may exaggerate symptom endorsement on self-report measures (e.g., Ovsiew et al., 2023; Sullivan et al., 2007). Feigning symptoms related to ADHD may also occur in the absence of external motive(s) like medications or accommodations.

The underlying reasons that compel both healthy and clinical populations to deliberately or inadvertently exaggerate, overestimate, or fabricate symptoms in the absence of secondary gain is not well-understood, and perhaps more pervasive than once believed (Dandachi-FitzGerald et al., 2020; Finley et al., 2023). Social media, however, may have a particularly strong influence on symptom overreporting in ADHD evaluations. Videos tagged #ADHD on the social media platform TikTok

CONTACT John-Christopher A. Finley jfinley3045@gmail.com Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School, 710 N Lake Shore Drive, Chicago, IL 60611, USA
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13803395.2023.2283940.

have amassed several billion views (Yeung et al., 2022). Yet, approximately half of these tagged videos provide misleading information about ADHD and another quarter provide nonspecific personal anecdotes of those with lived experience of ADHD (Yeung et al., 2022). Misleading information and/or misidentification of nonspecific experiences of inattention may prompt viewers to identify with a diagnosis of ADHD, seek further clinical evaluation, and use information gathered from TikTok to inform their responses in a clinical evaluation, potentially resulting in over-endorsement of symptoms on self-report questionnaires. This phenomenon is not uncommon given the burgeoning concern that viral TikTok trends may also lead to the misidentification and romanticization of mental health symptoms, and has even been shown to influence the manifestation of functional tic-like behaviors (Pringsheim et al., 2021).

Regardless of the underlying reason(s) for symptom overreporting, assessing the validity of self-reported symptoms is crucial to avoid inaccurate diagnosis and potentially harmful treatment recommendations (e.g., Harrison et al., 2007; Lakhan & Kirchgessner, 2012; Mitchell & Read, 2012). Symptom validity tests (SVTs) are helpful in discerning the veracity of self-reported symptoms and are typically embedded within self-report symptom rating inventories. Although the use of SVTs has gained much attention over the past several years, few studies have focused on cross-validating embedded SVTs specifically for ADHD evaluation (Tucha et al., 2015).

Most SVTs used in ADHD evaluations are embedded in personality inventories that assess the degree to which an examinee endorses an atypical number of symptoms (e.g., Minnesota Multiphasic Personality Inventory-2-Restructured Form [MMPI-2-RF]; Ben-Porath, 2012). Since these indicators rely on items that assess a wide range of psychopathology, they may not be as sensitive to indicators that exclusively index the validity of self-reported ADHD symptoms (e.g., Frazier et al., 2008; Jasinski et al., 2011; Marshall et al., 2010; Sollman et al., 2010). There is also a scarcity of embedded SVTs specifically validated to indicate the degree to which an examinee endorses an atypical number of ADHD symptoms that raise concern for invalid symptom endorsement. Further, given that some existing ADHD-specific validity indicators were established after the creation of the questionnaire from which they are derived (e.g., Conners Adult ADHD Rating Scale Infrequency Index; CAARS; Cook et al., 2016; Suhr et al., 2011), they are limited to items that assess common ADHD symptoms rather than highly unusual symptoms that may reflect feigning. The CAARS Exaggeration Index (Harrison & Armstrong, 2016) is among the few embedded validity indicators that assesses highly unusual symptoms related to ADHD. However, it requires adding five items that were not included in the original CAARS, which may necessitate additional materials and introduce complications to administration and scoring procedures, limiting its convenience for practitioners at this time. For most of the existing embedded validity indicators, conservative cut-scores may be required to ensure that the indicators can discriminate between individuals with genuinely severe symptoms from those feigning symptoms, which may consequently reduce the sensitivity.

One questionnaire that offers potentially sensitive indicators of feigned ADHD symptoms is the Clinical Assessment of Attention Deficit-Adult (CAT-A; Bracken & Boatwright, 2005). Unlike other popular ADHD symptom rating inventories, such as the CAARS (Conners et al., 1998) or Barkley Adult ADHD Rating Scale-4th Edition (Barkley, 2011), the CAT-A contains three symptom validity indicators, the Negative Impression (NI), Infrequency (IF), and Positive Impression (PI) scales, that were included as part of the original measure development and validation. Although the CAT-A clinical scales have demonstrated sufficient reliability and validity (Bracken & Boatwright, 2005), little is known about the psychometric properties of these SVTs. The cut-scores for these SVTs were derived by selecting values with a low base rate of endorsement in the validation sample (i.e., ≤5% of the normative sample) rather than cross-validation against an established criterion for independently determining invalid ADHD symptom reporting. Therefore, it remains unclear whether the recommended cut-scores in the manual optimize sensitivity and specificity for detection of invalid ADHD symptom endorsement.

Two recent studies have evaluated the convergent validity of the CAT-A validity indicators. Cullins and Park (2021) showed that the NI scale has moderate convergent validity with the Response Bias Scale (RBS), a well-validated embedded SVT from the MMPI-2-RF that assesses noncredible memory complaints. They also showed that the NI and IF scales have modest associations with the MMPI-2-RF Infrequent Responses (F-r) and Infrequent Psychopathology Response (Fp-r) scales, but it should be noted that their study sample consisted of nonclinical undergraduate volunteers. Leib et al. (2021) examined the three CAT-A SVTs and five MMPI-2-RF overreporting SVTs within a clinical adult ADHD sample, and similarly demonstrated moderate associations between each set of the

validity indicators, with a 51% overall concordance rate between the CAT-A and MMPI-2-RF SVTs. Most of this overall concordance rate arose from those with valid CAT-A and MMPI-2-RF responding (38%) as opposed to invalid responding across both measures (13%).

Based on the items comprising the CAT-A validity scales, it is possible that the IF scale may be the best indicator of symptom invalidity for the CAT-A. Marshall et al. (2010) conducted the only study cross-validating its utility for ADHD evaluations and reported that an IF cut-score of ⩾3 had adequate sensitivity (58%) and specificity (90%) in a clinical sample of adults with ADHD. However, their study used a combination of symptom and performance validity tests to establish criterion groups, making it difficult to interpret how sensitive a cut-score of ⩾3 is for detecting symptom invalidity. More recent studies have suggested that symptom and performance validity are dissociable within the context of ADHD and therefore should be evaluated independently (e.g., Sweet et al., 2021; White et al., 2022). Although symptom and performance validity are distinct constructs in ADHD assessment, these findings suggest the IF scale may also be sensitive to feigned neurocognitive performance, although further research is needed to substantiate this assertion. In particular, comparing how well the IF scale detects invalid symptom reporting versus neurocognitive performance would provide clinicians with a better sense of its construct validity and psychometric properties.

Current study

With the increase in ADHD diagnoses and referrals for neuropsychological evaluations (Fayyad et al., 2017; Schroeder et al., 2019), there is a need to develop more SVTs that can be used in ADHD evaluations. Although the CAT-A offers promising SVTs for ADHD evaluations, there is minimal research to demonstrate the specific psychometrics of these indicators in clinical practice. To our knowledge, no study has cross-validated all three CAT-A symptom validity scales in a single ADHD sample. Therefore, this study investigated how well the CAT-A NI, IF, and PI validity scales could classify the validity of symptom overreporting and underreporting and neurocognitive performance during a ADHD evaluation. This study also aimed to identify optimal cut-scores for each of these scales and compare how well they perform individually and in combination among a large sample of adults undergoing an ADHD evaluation.

Method

Participants and procedures

This retrospective cross-sectional study comprised 599 consecutive adult outpatients who underwent a comprehensive neuropsychological evaluation at an academic medical center to aid in differential diagnosis and treatment planning related to known or suspected ADHD. All patients provided written informed consent for their data to be included in a larger, ongoing IRB-approved research study between the years 2018 and 2023. Patients were excluded ($n = 50$) if they had missing data. Participants were also excluded ($n = 6$) if they had invalid inconsistency scores on the MMPI-2-RF (T-scores ≥80 on the True and Variable Response Inconsistency Scales; Ben-Porath & Tellegen, 2008) to ensure their reporting of symptoms was deliberate and not confounded by inconsistent or careless responding. No participants omitted more than 2 items on the MMPI-2-RF. The final sample comprised 543 patients after applying the exclusion criteria. Approximately 30% were diagnosed with ADHD only ($n = 161$), 48% with ADHD and a comorbid psychiatric disorder ($n = 263$), 17% with only a primary psychiatric disorder ($n = 94$), typically including depression, anxiety, or posttraumatic stress disorder, 2% with no diagnosis ($n = 13$), and 2% with another mental health disorder ($n = 12$). See Table 1 for demographics and clinical characteristics for the entire sample and validity groupings.

Patients were divided into three separate validity groups based on scores from multiple SVTs and performance validity tests (PVTs) that assess constructs similar to the CAT-A validity indicators (see the Measures section below). Because the CAT-A NI and IF scales are designed to assess symptom overreporting, we examined how well these scales discriminate valid and invalid groups based on measures of overreporting, with the latter being defined by ⩾2 failures on measures of overreporting. To expand upon Marshall et al. (2010) study, we also examined how well the NI and IF could discriminate valid versus invalid groups defined by ⩾2 failures on measures of exaggerated neurocognitive performance. Because the PI scale is intended to assess symptom underreporting, we examined how well it discriminated valid and invalid groups based on measures of underreporting, with the latter defined by ⩾1 failure on measures of underreporting. We used conservative cutoffs to identify patients with invalid symptom underreporting because (1) few patients in our sample failed more than one of the underreporting criterion measures and (2) the criterion SVTs for this group were

Table 1. Sample characteristics and descriptive statistics.

Demographics and Clinical Characteristics	Total Sample ($N = 543$)	Symptom Overreporting Group ($n = 128$)	Symptom Underreporting Group ($n = 15$)	Invalid Performance Group ($n = 48$)
Means & Proportions				
Age (years)	$M = 28.13$ ($SD = 7.13$)	$M = 26.22$ ($SD = 5.95$)	$M = 25.98$ ($SD = 4.89$)	$M = 26.52$ ($SD = 6.74$)
Education (years)	$M = 15.83$ ($SD = 2.06$)	$M = 15.03$ ($SD = 2.09$)	$M = 14.93$ ($SD = 2.06$)	$M = 14.56$ ($SD = 2.23$)
Sex: Female	341 (62%)	92 (72%)	7 (47%)	22 (49%)
Racial Identity				
Non-Hispanic White	248 (45%)	44 (34%)	4 (27%)	14 (29%)
Hispanic	122 (22%)	30 (23%)	4 (27%)	16 (33%)
Non-Hispanic Black	91 (17%)	28 (22%)	5 (33%)	14 (29%)
Asian/Pacific Islander	55 (10%)	15 (12%)	1 (7%)	3 (6%)
Multiracial	32 (6%)	11 (9%)	1 (7%)	1 (2%)
Handedness				
Right	518 (95%)	117 (91%)	13 (87%)	43 (90%)
Left	30 (5%)	11 (9%)	2 (13%)	5 (10%)
Ambidextrous	1 (<1%)	0 (0%)	0 (0%)	0 (0%)
Student Status				
Actively enrolled	378 (69%)	88 (69%)	13 (87%)	31 (65%)
Non-student	171 (31%)	40 (31%)	2 (13%)	17 (35%)
Estimated Premorbid IQ	$M = 106.08$ ($SD = 7.88$)	$M = 104.00$ ($SD = 7.62$)	$M = 103.15$ ($SD = 7.52$)	$M = 103.15$ ($SD = 7.09$)
CAT-A SVTs				
NI Atypical	89 (16%)	49 (38%)	1 (7%)	11 (23%)
NI Very Atypical	8 (1%)	6 (5%)	0 (0%)	2 (4%)
IF Atypical	40 (7%)	20 (16%)	1 (7%)	5 (10%)
IF Very Atypical	11 (2%)	4 (3%)	0 (0%)	1 (2%)
PI Atypical	5 (1%)	0 (0%)	5 (33%)	1 (2%)
PI Very Atypical	3 (1%)	0 (0%)	1 (7%)	0 (0%)

Note: M = mean; SD = standard deviation; Estimated Premorbid IQ = Estimated Premorbid IQ based on demographically-adjusted standard scores from Test of Premorbid Functioning; CAT-A SVTs = Clinical Assessment of Attention Deficit-Adult symptom validity tests; NI = Negative Impression; IF = Infrequency; PI = Positive Impression.

embedded within a single questionnaire and generally intercorrelated. See Table 2 for more information regarding criterion grouping.

Measures

CAT-A symptom validity tests

The CAT-A is a self-report questionnaire that assesses childhood and adulthood ADHD symptomatology aligned with diagnostic criteria in the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition-Text Revision (American Psychiatric Association, 2000; Bracken & Boatwright, 2005). It comprises 108 items rated on a four-point Likert scale, from "Strongly Disagree" to "Strongly Agree." Selecting "Strongly Agree" and "Strongly Disagree" is considered a "negative" and "positive" symptom endorsement, respectively. Negative and positive symptom endorsements are descriptors reflecting the most severe symptom ratings that are used to define the CAT-A validity indicator scores.

Specifically, the sum number of negative or positive symptom endorsements among participants in the original normative sample were used to create the cut-scores for the CAT-A embedded validity indicators. As discussed above, the NI scale assesses the severity of negative symptom endorsement, with scores 46–77 indicating "Atypical" and 78–108 indicating "Very Atypical" symptom overreporting. The IF scale also assesses the severity of negative symptom endorsement using 10 items that were endorsed by less than 2% of the original normative sample. Scores 4–5 indicate "Atypical" and 6–10 indicate "Very Atypical" symptom overreporting. The manual indicates that elevated scores on the IF scale are most indicative of symptom exaggeration, although they may reflect genuinely severe symptoms in some individuals. The PI scale assesses positive symptom endorsement, with scores 63–79 indicating "Atypical" and 80–108 indicating "Very Atypical" symptom underreporting. No psychometric properties regarding the classification accuracy of these scales were reported in the CAT-A manual.

Criterion symptom validity tests

Empirically established embedded SVTs from the MMPI-2-RF (Ben-Porath & Tellegen, 2008) were used as the criterion measures to establish the overreporting validity groups. Specifically, we used the following

Table 2. Criterion symptom and performance validity tests.

Criterion Validity Tests	Cutoff (Metric)	Failure Cutoff Reference(s)	Sample Failure Rate
Criterion Symptom Validity Tests for Overreporting			
MMPI-2-RF F-r	≥83 (T-score)	Morris et al. (2022)	142/543 (26%)
MMPI-2-RF Fp-r	≥85 (T-score)	Harp et al. (2011), Robinson and Rogers (2018)	116/543 (21%)
MMPI-2-RF Fs	≥91 (T-score)	Harp et al. (2011), Robinson and Rogers (2018)	72/543 (13%)
MMPI-2-RF RBS	≥80 (T-score)	Morris et al. (2022)	148/543 (27%)
Criterion Performance Validity Tests for Exaggerated Neurocognitive Performance			
BVMT-R Recognition Discrimination	≤4 (raw)	Bailey et al. (2018), Phillips et al. (2023)	15/543 (3%)
Dot Counting Test E-Score	≥14 (raw)	Abramson et al. (2021)	53/543 (10%)
Reliable Digit Span	≤7 (raw)	Bing-Canar et al. (2022)	54/543 (10%)
Rey 15-Item Test Recall/Recognition	≤23 (raw)	Ashendorf et al. (2021)	26/543 (5%)
RAVLT Effort Score	≤13 (raw)	Tse et al. (2023)	82/543 (15%)
Criterion Symptom Validity Tests for Underreporting			
MMPI-2-RF L-r	≥81 (T-score)	Ben-Porath and Tellegen (2008)	14/543 (3%)
MMPI-2-RF K-r	≥81 (T-score)	Ben-Porath and Tellegen (2008)	1/543 (<1%)

Note: $N = 543$; MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias Scale; L-r = Uncommon Virtues; K-r = Adjustment Validity; BVMT-R = Brief Visuospatial Memory Test-Revised; RAVLT = Rey Auditory Verbal Learning Test.

SVTs: Infrequent Responses (F-r); Infrequent Psychopathology Responses (Fp-r); Infrequent Somatic Responses (Fs); Response Bias Scale (RBS). The cut-scores for these SVTs that were used to develop the overreporting groups were based on ADHD cross-validation studies (Harp et al., 2011; Morris et al., 2022; Robinson & Rogers, 2018). We chose the most conservative cut-scores from these studies since they were based on simulation designs, which often yield more overly optimistic findings compared to known groups designs (Rogers, 1997). Nonetheless, the cut-scores from these studies were far more liberal than those proposed in the original MMPI-2-RF manual (Ben-Porath & Tellegen, 2008). Given these discrepancies, we analyzed the data with cut-scores based on ADHD cross-validation studies as well as the MMPI-2-RF manual. The findings reported in this paper were based on the cross-validation cut-scores, and the results based on the MMPI-2-RF manual cut-scores are reported in Supplementary Table S1.

The Uncommon Virtues (L-r) and Adjustment Validity (K-r) embedded SVTs from the MMPI-2-RF were used to establish the underreporting validity groups. However, these cut-scores were based exclusively on the MMPI-2-RF manual since we could not find any empirical research proposing adjusted cutoffs for these indices in ADHD populations. See Table 2 for the exact cut-scores and number of patients who were deemed to have invalid criterion SVTs.

Criterion performance validity tests

Five empirically established freestanding and embedded PVTs with cutoffs that have previously been cross-validated in adult ADHD samples were used as the criterion measures to establish the neurocognitive performance validity groups. The freestanding criterion PVTs included the Dot Counting Test E-Score (Boone et al., 2002) and Rey 15-Item Test Recall + Recognition (Boone et al., 2002). The embedded criterion PVTs included the Brief Visuospatial Memory Test-Revised Recognition Discrimination (Bailey et al., 2018), Reliable Digit Span (Greiffenstein et al., 1994), and Rey Auditory Verbal Learning Test Effort Score (Boone et al., 2005). See Table 2 for the specific cut-scores that were used in this study and the number of patients who were deemed to have invalid criterion PVTs.

Data analyses

Frequency analyses were conducted after patients were classified into validity groupings based on criterion measures of symptom over/underreporting and exaggerated neurocognitive performance. These analyses were conducted to determine the rates at which adults undergoing ADHD evaluations score below the cutoff on multiple SVTs and PVTs.

Given the skewed nature of the CAT-A validity indicators, nonparametric tests were used for some analyses. For instances in which parametric analyses were needed, a square root transformation was used to normalize the score distribution. Post-hoc power analyses indicated that all findings had an observed power greater than 80%. Preliminary analyses included Mann-Whitney U Tests to examine potential differences in CAT-A validity scores between the valid and invalid groups, and Spearman correlations to examine relationships and redundancy among the scores on the CAT-A validity scales and the criterion SVTs for those in the valid groups.

For the primary analyses, receiver operating characteristic (ROC) curve analyses were conducted to derive cut-scores from the CAT-A validity indicators that could significantly discriminate invalid from valid groups, maximizing sensitivity while maintaining ⩾.90 specificity. Cut-scores with these specificity values were required to ensure no more than 10% of patients with valid data were misclassified as having invalid data, which is a well-accepted recommendation for clinical and research practices (Nitch & Glassmire, 2007). Sensitivity and specificity rates were used to calculate the positive and negative predictive accuracy for identifying invalid symptom and performance validity among each group in our sample. The overall classification accuracy of each validity indicator was based on area under the curve values, with poor accuracy values ranging from .50 to .69, acceptable ranging from .70 to .79, excellent ranging from .80 to .89, and outstanding ranging from .90 to 1.00 (Hosmer et al., 2013).

Simple logistic regression analyses were then conducted to examine how much variance in symptom validity status the CAT-A validity scales explained in isolation. Because the NI and IF scales are both designed to measure overreporting, a multivariate logistic regression analysis was conducted to examine whether there was any incremental value in using these SVTs together versus independently.

Results

Sample base rates of invalid symptom reporting

Approximately 24% of the sample ($n = 128$) produced invalid scores on the criterion measures of symptom overreporting, including the MMPI-2-RF F-r, Fp-r, Fs, and RBS. Approximately 3% of the sample ($n = 15$) produced invalid scores on the criterion measures of symptom underreporting, including the MMPI-2-RF L-r and K-r. By comparison, 9% ($n = 48$) produced invalid scores on two or more of the criterion measures of invalid neurocognitive performance, including either the Dot Counting Test E-Score, Rey 15-Item Test Recall + Recognition, Brief Visuospatial Memory Test-Revised Recognition Discrimination, Reliable Digit Span, or Rey Auditory Verbal Learning Test Effort Score. See Table 2 for the failure rates per criterion measure.

Group differences

Mann-Whitney U tests showed that the invalid group ($M = 38.97$; $SD = 17.18$), defined by symptom *overreporting* criterion measures, had significantly higher NI scores than the valid group ($M = 27.07$; $SD = 16.54$), with medium effects ($W = 1696.00$, $p < .001$, $r = .39$). Similarly, the invalid group ($M = 2.11$; $SD = 1.65$), defined by symptom *overreporting* criterion measures, had significantly higher IF scores than the valid group ($M = 1.36$; $SD = 1.51$), with medium effects ($W = 2153.00$, $p = .001$, $r = .20$); but there was no significant difference ($p > .05$) in PI scores. The invalid group ($M = 26.93$; $SD = 26.66$), when defined by symptom *underreporting* criterion measures, had significantly higher PI scores than the valid group ($M = 11.91$; $SD = 10.90$) with a medium effect ($W = 2435.50$, $p = .002$, $r = .34$) and no significant differences in NI and IF scores. When defined by performance validity criterion measures, the invalid group had significantly higher scores on the NI ($M = 38.96$; $SD = 14.99$; $W = 7506.00$, $p < .001$, $r = .37$) and IF ($M = 1.83$; $SD = 1.58$; $W = 9762.50$, $p = .035$, $r = .18$) than the valid groups' NI ($M = 28.74$; $SD = 1.73$) and IF scores ($M = 1.44$; $SD = 1.55$), with a medium and small effect, respectively. The invalid group ($M = 8.19$; $SD = 7.79$), when defined by performance validity criterion measures, had significantly lower PI scores ($W = 14478.00$, $p = .012$, $r = .22$) than the valid group ($M = 12.70$; $SD = 12.01$), with a small effect. No statistically significant demographic differences were found between the invalid-valid groupings when defined by overreporting, underreporting, or exaggerated performance measures (see Table 1 for details).

Table 3. Symptom validity tests intercorrelations for participants with valid symptom reporting and neurocognitive performance.

	IF	NI	PI	F-r	Fp-r	Fs	RBS	L-r	K-r
IF	–	.77***	−.18***	.20***	.14**	.18***	−.02	−.02	−.05
NI		–	−.32***	.36***	.27***	.32***	.10*	−.10*	−.17***
PI			–	−.10*	−.06	−.14**	.02	.27***	.14**
F-r				–	.52***	.63***	.64***	−.23***	−.57***
Fp-r					–	.38***	.36***	−.17***	−.40***
Fs						–	.50***	−.17***	−.39***
RBS							–	−.03	−.41***
L-r								–	.37***
K-r									–

Note: $n = 374$; IF = Infrequency; NI = Negative Impression; PI = Positive Impression; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; RBS = Response Bias Scale; L-r = Uncommon Virtues; K-r = Adjustment Validity.
*$p < .05$; **$p < .01$; ***$p < .001$.

Correlations among symptom validity tests

As shown in Table 3, the SVTs embedded within the CAT-A and MMPI-2-RF (i.e., the criterion SVTs) demonstrated a wide range of correlations with each other. Nonetheless, the validity indicators generally correlated in the expected directions based on the constructs they are intended to measure. For instance, negative correlations were found between overreporting and underreporting validity scales. The NI scale correlated most with the MMPI-2-RF overreporting scales. Weaker correlations were observed between the IF scale and MMPI-2-RF overreporting scales. The PI scale showed modest correlations with the MMPI-2-RF underreporting scales. When examining the intercorrelations among the CAT-A validity scales, the IF and NI scales were strongly correlated. Unsurprisingly, the PI scale was negatively and modestly correlated with the IF and NI scales.

Negative impression cutoffs and classification accuracy statistics

ROC curve analyses revealed that NI scores significantly differentiated valid from invalid symptom overreporting and neurocognitive performance (see Table 4). The NI scale generally demonstrated acceptable classification accuracies of .70 and .69 when predicting validity status in the groups defined by symptom overreporting and neurocognitive performance measures, respectively. A raw score of ⩾51 and ⩾53 was the optimal cutoff for differentiating invalid from valid symptom reporting and neurocognitive performance, respectively. Less than 10% of participants with valid data were misclassified as having invalid data based on these cut-scores; however, the NI cut-score was more sensitive to invalid symptom reporting (.30) than invalid neurocognitive performance (.15).

Infrequency cutoffs and classification accuracy statistics

Similar results were found when identifying optimal cut-scores for the IF scale (see Table 5). ROC curve analyses revealed that IF scores significantly differentiated valid from invalid symptom overreporting and neurocognitive performance. However, the IF demonstrated poor classification accuracies of .64 and .58. The same optimal cut-score (raw score of ⩾4) was shown to classify both invalid symptom reporting and invalid neurocognitive performance. Less than 10% of patients with valid data were misclassified as having invalid data based on this cut-score; however, this cutoff was also relatively more sensitive to invalid

Table 4. Classification accuracy and optimal CAT-A negative Impression cut-scores for classifying invalid symptom overreporting and neurocognitive performance.

Invalid Criterion Method	AUC (95% CI)	Cutoff	SN	SP	10% Base Rate		20% Base Rate		30% Base Rate	
					PPV	NPV	PPV	NPV	PPV	NPV
Invalid Symptom Validity Group		≥52	.24	.92	.25	.92	.43	.83	.56	.74
(n = 128)	.70*** (.65, .75)	≥51	.30	.90	.25	.92	.43	.84	.56	.75
		≥50	.33	.89	.25	.92	.43	.84	.56	.76
Invalid Performance Validity Group (n = 48)		≥54	.14	.91	.15	.90	.28	.81	.40	.71
	.69*** (.63, .76)	≥53	.15	.90	.14	.91	.27	.81	.39	.71
		≥52	.17	.89	.15	.91	.28	.81	.40	.71

Note: AUC = Area Under the Curve; CI = Confidence Interval; SN = Sensitivity; SP = Specificity; PPV = Positive Predictive Power; NPV = Negative Predictive Power; CAT-A = Clinical Assessment of Attention Deficit-Adult.
*p < .05; **p < .01; ***p < .001.

Table 5. Classification accuracy and optimal CAT-A Infrequency cut-scores for classifying invalid symptom overreporting and neurocognitive performance.

Invalid Criterion Method	AUC (95% CI)	Cutoff	SN	SP	10% Base Rate		20% Base Rate		30% Base Rate	
					PPV	NPV	PPV	NPV	PPV	NPV
Invalid Symptom Validity Group (n = 128)		≥5	.05	.96	.12	.90	.24	.80	.35	.70
	.64*** (.54, .74)	≥4	.18	.90	.17	.91	.31	.81	.44	.72
		≥3	.39	.71	.13	.91	.25	.82	.37	.73
Invalid Performance Validity Group (n = 48)		≥5	.05	.95	.10	.90	.20	.80	.30	.70
	.58** (.51, .65)	≥4	.09	.90	.09	.90	.18	.80	.28	.70
		≥3	.29	.73	.11	.90	.21	.80	.32	.71

Note: AUC = Area Under the Curve; CI = Confidence Interval; SN = Sensitivity; SP = Specificity; PPV = Positive Predictive Power; NPV = Negative Predictive Power; CAT-A = Clinical Assessment of Attention Deficit-Adult.
*p < .05; **p < .01; ***p < .001.

Table 6. Classification accuracy and optimal CAT-A positive Impression cut-scores for classifying invalid symptom underreporting.

Invalid Criterion Method	AUC (95% CI)	Cutoff	SN	SP	10% Base Rate		20% Base Rate		30% Base Rate	
					PPV	NPV	PPV	NPV	PPV	NPV
Invalid Symptom Validity Group ($n = 15$)	.70*** (.51, .84)	≥28	.35	.91	.30	.93	.49	.85	.63	.77
		≥27	.36	.90	.29	.93	.47	.85	.61	.77
		≥26	.36	.88	.25	.93	.43	.85	.56	.93

Note: AUC = Area Under the Curve; CI = Confidence Interval; SN = Sensitivity; SP = Specificity; PPV = Positive Predictive Power; NPV = Negative Predictive Power; CAT-A = Clinical Assessment of Attention Deficit-Adult.
*$p < .05$; **$p < .01$; ***$p < .001$.

symptom reporting (.18) than invalid neurocognitive performance (.09).

Positive impression cutoffs and classification accuracy statistics

ROC curve analyses revealed that PI scores significantly discriminated valid from invalid symptom underreporting, with an acceptable classification accuracy of .70 (see Table 6). When ensuring ⩾.90 specificity, the optimal PI cut-score (raw score of ⩾27) yielded better sensitivity (.36) than the other CAT-A SVTs.

Univariate and multivariate predictive models of the symptom validity tests

Simple logistic regression analyses indicated that NI ($\chi 2 = 41.73$, $\beta = 1.52$, $p < .001$, 95% CI [1.33, 1.75]) and IF scores ($\chi 2 = 11.12$, $\beta = 1.58$, $p = .001$, 95% CI [1.21, 2.09]) significantly predicted group membership based on patients who failed the criterion validity measures of symptom overreporting. The NI scale explained a larger proportion of variance in validity status (Nagelkerke's $R^2 = .11$) than the IF scale (Nagelkerke's $R^2 = .04$). Interestingly, the ability to predict group membership did not significantly improve when combining the NI and IF scores into a model together as compared to examining them independently. The multivariate logistic regression model explained the same amount of variance (Nagelkerke's $R^2 = .11$) as the NI scale on its own, and the IF did not significantly contribute to the multivariate model, unlike the NI, suggesting the IF scale indexed redundant aspects of the variance in validity status. Although the IF and NI variables were strongly correlated, they did not violate the assumption of multicollinearity for the multivariate regression. The PI scores significantly predicted group membership based on those who failed one or more of the criterion validity measures of symptom underreporting ($\chi 2 = 9.76$, $\beta = 1.63$, $p = .002$, 95% CI [1.20, 2.21]), but explained a modest amount of the systematic variance (Nagelkerke's $R^2 = .08$) in validity status.

Discussion

The current study investigated the CAT-A validity indicators of symptom overreporting and underreporting among a large sample of adults who presented for an ADHD evaluation. Approximately 27% and 9% of the patients within this sample produced invalid scores on multiple empirical symptom and performance validity tests, respectively. However, the base rate of invalid symptom reporting varied depending on which criterion measures were used to classify validity status. Many fewer participants failed measures of symptom underreporting (3%) than symptom overreporting (24%). The failure rate of performance validity in this study generally aligned, though it was on the lower end of the spectrum, with previous research using multiple criterion measures to define validity status (Hirsch et al., 2022; Martin & Schroeder, 2020; Ovsiew et al., 2023; Phillips et al., 2023). The failure rates of overreporting in this study were consistent with prior research (Ovsiew et al., 2023; Sullivan et al., 2007). Although little is known about the base rates of underreporting in ADHD evaluations, the incidence of underreporting is not unexpected given this is a clinically presenting sample primarily referred on the basis of ADHD complaints. These findings illustrate the importance of separating underreporting and overreporting criterion measures to classify symptom validity status when using a known-groups design. Cultivating separate validity groupings on this basis was crucial for the current study since the CAT-A IF and NI scales are intended to assess overreporting, whereas the PI scale is meant to assess underreporting.

The current study findings demonstrated that the CAT-A underreporting and overreporting validity indicators capture dissociable constructs. Scores on the overreporting NI and IF scales were highly correlated and negatively associated with scores on the PI underreporting scale, providing evidence for convergence validity. Modest correlations between the CAT-A and MMPI-2-RF validity scales indicated questionable concurrent validity; however, the overreporting and

underreporting scales from the CAT-A were exclusively (though, modestly) associated with overreporting and underreporting scales from the MMPI-2-RF, respectively. The weak correlations between the CAT-A and MMPI-2-RF validity scales may be explained by the content, rather than severity, of the symptoms that they assess; the CAT-A assesses symptoms of ADHD, whereas the MMPI-2-RF assesses broad symptoms of psychopathology. The concept that the IF and NI scales assess overreporting and the PI scale assesses underreporting is generally supported by these preliminary findings.

Group comparison analyses expanded upon the psychometric properties of the CAT-A indicators. Elevated IF/NI and PI scores were significantly higher among adults with invalid scores on empirical measures of overreporting and underreporting, respectively. Although the average CAT-A validity scores significantly predicted validity status, they only explained a modest proportion of variance (4%–11%) in validity membership. The ability to predict validity status did not improve when using the NI and IF scores in a combined model, with the NI showing more incremental value. These multivariate and univariate analyses in the context of prior research findings (i.e., Cullins & Park, 2021; Leib et al., 2021), suggest that chaining IF and NI scores may be redundant and may not improve the detection of invalid symptom responding. These findings make sense given that part of the NI scale is composed of the same items as the IF. Thus, elevated IF scores will also elevate NI scores, but not necessarily the other way around. Group comparison analyses also showed that elevated IF/NI and PI scores were moderately positively and negatively associated with exaggerated neurocognitive performance, respectively. However, the strength of association between scores on the CAT-A scales and invalid neurocognitive performance was minimal when we explored the classification accuracy statistics. These two latter findings are generally consistent with prior research (White et al., 2022) suggesting that in the context of an ADHD evaluation, scores on symptom and performance validity tests reflect similar, but dissociable constructs.

Only one other study (i.e., Marshall et al., 2010) has examined the classification accuracy statistics of one CAT-A validity scale. To our knowledge, the current study is the first to investigate which cut-scores in all three CAT-A validity indicators may be used to index the validity of symptom reporting and neurocognitive performance in adults undergoing an ADHD evaluation. This research is also among the few extant studies cross-validating symptom validity indicators that are embedded within an ADHD questionnaire.

Findings revealed that cut-scores on the NI and IF scales could be used as embedded SVTs of overreporting. Scores $\geqslant$51 on the NI scale adequately discriminated between adults with and without invalid scores on empirical SVTs of overreporting. This cut-score yielded sufficient specificity (≥90%) and modest sensitivity (30%), adding clarity to the manual-specified ranges of 46–77 for "Atypical" and 78–108 for "Very Atypical" responding (Bracken & Boatwright, 2005). Scores $\geqslant$4 on the IF scale significantly, but poorly discriminated between adults with and without invalid scores on empirical SVTs of overreporting. This cut-score is consistent with the manual-specified cutoff for "Atypical" responding (Bracken & Boatwright, 2005). Marshall et al. (2010) found that a cutoff of $\geqslant$3 was optimal for the IF scale. However, their findings revealed a higher sensitivity value (58%) than observed in our sample (39%) when using a cut-score of $\geqslant$3. The discrepancy between these sensitivity values may be influenced by differences in the criterion measures used to determine the validity groupings between studies.

Classification accuracy statistics were attenuated when examining how well the NI and IF scales discriminated patients with and without invalid neurocognitive performance. Optimal cut-scores (≥53 for NI and $\geqslant$4 for IF) were similar to those discovered in the prior analyses, but sensitivity values were unacceptably low (15% and 9%, respectively) when the cut-scores were adjusted for a specificity of $\geqslant$90% (Power et al., 2013). These results further support extant research demonstrating that symptom and performance validity tests should not be interpreted interchangeably (e.g., Sweet et al., 2021; White et al., 2022).

We also identified a cut-score on the PI scale that could be used as an embedded SVT of underreporting. Specifically, PI scores $\geqslant$27 adequately discriminated between adults with and without invalid scores on empirical SVTs of underreporting. This cut-score also yielded adequate specificity and modest sensitivity (36%). Interestingly, this cut-score is far more conservative than the manual-specified cutoff of 63-79 for "Atypical" responding. However, the CAT-A cutoff descriptors ("Atypical" and "Very Atypical") used in the manual are not necessarily intended to be used interchangeably with symptom validity classification (Bracken & Boatwright, 2005).

As with many embedded SVTs, the identified cut-scores in this study yielded low sensitivity values. None of these cut-scores upheld the "Larrabee limit" (i.e., ≥.50 sensitivity and ≥ .90 specificity; Larrabee, personal communication, November 2007), indicating that they should not be used in isolation to determine the validity of symptom endorsement. In addition to

the sensitivity and specificity values, practitioners should also consider the positive and negative predictive values of the NI, IF, and PI. As indicated in Tables 4, 5 , and 6, the optimal IF (≥4), NI (≥51), and PI (≥27) cutoffs yield low positive predictive values, suggesting that practitioners would have a 17%, 25%, and 29% probability, respectively, of accurately identifying someone with invalid symptom reporting at a 10% invalidity base rate. Conversely, the negative predictive values suggest that there is a 91–93% probability of correctly not suspecting someone of invalid symptom reporting if they do not fail these CAT-A cutoffs. Even if the base rate of invalid symptom reporting hypothetically increased to 20% or 30%, the positive predictive values for these cutoffs would remain low. The negative predictive values at these base rates would also diminish, but would remain much higher than the positive predictive values. In short, these positive and negative predictive values further suggest that the CAT-A NI, IF, and PI cutoffs may fail to detect a large percentage of individuals with invalid symptom ADHD reporting in most clinical settings. Practitioners may therefore wish to interpret scores below these cutoffs within the context of other empirical validity measures as well as the patient's behavioral presentation and clinical and medical history.

The current study is not without limitations. It is first important to acknowledge that the criterion validity tests used in this study are not perfect indicators of performance and symptom validity status. While the symptom invalidity rate in this study was consistent with prior research (see Ovsiew et al., 2023), it is possible that relying on embedded SVTs from a broad psychopathology questionnaire (MMPI-2-RF) may have limited our ability to detect individuals who feigned symptoms related to ADHD. Adults feigning a diagnosis of ADHD may strategically exaggerate symptoms that are only associated with ADHD (Marshall et al., 2010). Thus, including ADHD-specific SVTs, such as the CAARS Infrequency Index (Cook et al., 2016) or Exaggeration Index (Harrison & Armstrong, 2016), as the criterion measures may have yielded higher and more accurate base rates of symptom invalidity in our sample.

The base rate of invalid performance in this study was also slightly lower than expected based on prior research (see Hirsch et al., 2022). Although we systematically chose criterion PVTs with diverse detection paradigms that have been cross-validated in ADHD populations (e.g., Abramson et al., 2021; Ashendorf et al., 2021; Bailey et al., 2018; Bing-Canar et al., 2022; Phillips et al., 2023; Tse et al., 2023), some may have been less sensitive to invalid performance than others. The Rey 15-Item Test and Brief Visuospatial Memory Test-Revised Recognition Discrimination criterion measures yielded particularly low failure rates in our sample. These PVTs may be robust in clinical populations with genuine and significant neuropsychiatric impairment, but insensitive to a sample like ours comprising highly educated adults with minimal cognitive impairment. It is important for future researchers to consider how certain PVTs, like the Rey 15-Item Test, may serve better at ruling in rather than ruling out invalid performance (Nitch & Glassmire, 2007). Future researchers may also wish to further explore which *a priori* statistical methods would help discern a combination of PVTs that is best suited for defining validity status in an ADHD sample (for review, see Finley et al., 2023).

It is also important to note that only one failure among the criterion measures for underreporting was used to determine validity status in one of the groupings. Although the PI demonstrated acceptable concurrent validity with the MMPI-2-RF underreporting scales, its generalizability as a validity index for underreporting is limited by using this lenient criterion. As such, the identified cut-score for this measure should be interpreted with caution, especially since it was highly inconsistent with the cut-scores reported in the CAT-A manual. Future research should use two or more failures on criterion measures of symptom underreporting to determine validity status. Another similar limitation of the current study was using only the embedded SVTs from the MMPI-2-RF to establish the validity groups for underreporting and overreporting. While the MMPI-2-RF comprises multiple validity indicators designed to assess non-redundant aspects of symptom reporting that have been cross-validated in ADHD samples, the scales are within a singular measure and were generally intercorrelated. We attempted to reconcile this issue by using cutoffs that have demonstrated high sensitivity and specificity in ADHD studies (Harp et al., 2011; Morris et al., 2022; Robinson & Rogers, 2018) to determine the validity groupings. Nonetheless, the likelihood of detecting invalid responding is diminished when using highly correlated SVTs (Rosenfeld et al., 2000), which may have influenced the low sensitivity values found in this study. Future research should use non-overlapping SVTs as criterion measures to determine validity status.

These findings provide preliminary evidence that cut-scores (≥4, ≥51, and ⩾27, respectively) from the CAT-A IF, NI, and PI scales may be used to assess the validity of symptom underreporting and overreporting in adults who are referred for an ADHD evaluation. However, using IF and NI scales together may not

improve the ability to detect overreporting of symptoms. Given the low sensitivity values, especially for the IF scale, and paucity of cross-validation research for these validity indicators, we encourage practitioners to use caution before relying on these proposed cut-scores during a neuropsychological evaluation. These CAT-A scales may help practitioners determine the veracity of symptom endorsement, but should not be used independently to classify symptom validity status. We also discourage the use of these symptom validity scales to guide clinical judgment regarding the validity of neurocognitive performance. We hope these findings spur further research into cross-validating the CAT-A scales as SVTs for ADHD evaluations.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Jason R. Soble http://orcid.org/0000-0003-3348-8762

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*, *30*(2), 264–273. https://doi.org/10.1177/10731911211050895

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders-fourth edition-text revision*.

Ashendorf, L., Clark, E. L., & Humphreys, C. T. (2021). The Rey 15-item memory test in US veterans. *Journal of Clinical and Experimental Neuropsychology*, *43*(3), 324–331. https://doi.org/10.1080/13803395.2021.1932761

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

Barkley, R. A. (2011). *Barkley adult ADHD Rating scale-IV (BAARS-IV)*. Guilford Press.

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

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*(5), 678–688. https://doi.org/10.1177/07342829221081921

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

Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002). The Rey 15-item recognition trial: A technique to enhance sensitivity of the Rey 15-item memorization test. *Journal of Clinical and Experimental Neuropsychology*, *24*(5), 561–573. https://doi.org/10.1076/jcen.24.5.561.1004

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

Conners, C. K., Erhardt, D., & Sparrow, E. P. (1998). *Conners adult attention Rating scale–self-Report: Long version*. Multi-Health Systems.

Cook, C. M., Bolinger, E., & Suhr, J. (2016). Further validation of the conner's adult attention deficit/hyperactivity Rating scale infrequency index (CII) for detection of non-credible report of attention deficit/hyperactivity disorder symptoms. *Archives of Clinical Neuropsychology*, *31*(4), 358–364. https://doi.org/10.1093/arclin/acw015

Cullins, B., & Park, S. (2021). Analysis of the Clinical Assessment of Attention Deficit-Adult in Comparison to the Minnesota Multiphasic Personality Inventory 2 Restructured Form in an Adult Psychoeducational Clinic. *The Undergraduate Journal of Psychology*, *32*(1), 40–50.

Dandachi-FitzGerald, B., Merckelbach, H., Bošković, I., & Jelicic, M. (2020). Do you know people who feign? Proxy respondents about feigned symptoms. *Psychological Injury and Law*, *13*(3), 225–234. https://doi.org/10.1007/s12207-020-09387-6

Fayyad, J., Sampson, N. A., Hwang, I., Adamowski, T., Aguilar-Gaxiola, S., Al-Hamzawi, A., Andrade, L. H. S. G., Borges, G., de Girolamo, G., Floresycu, S., Gureje, O., Haro, J. M., Hu, C., Karam, E. G., Lee, S., Navarro-Mateu, F., O'Neill, S., Pennell, B.-E., Piazza, M., & Posado-Villa, J., WHO World Mental Health Survey Collaborators. (2017). The descriptive epidemiology of DSM-IV adult ADHD in the world health organization world mental health surveys. *ADHD Attention Deficit & Hyperactivity Disorders*, *9*(1), 47–65.

Finley, J. A. Brooks, J. M. Nili, A. N. Oh, A. VanLandingham, H. B. Ovsiew, G. P. Ulrich, D. M. Resch, Z. J. & Soble, J. R. (2023). Multivariate examination of embedded indicators of performance validity for ADHD evaluations: A targeted approach. *Applied Neuropsychology. Adult*, 1–14. https://doi.org/10.1080/23279095.2023.2256440

Finley, J. C. A., Cladek, A., Gonzalez, C., & Brook, M. (2023). Perceived cognitive impairment is related to internalizing psychopathology but unrelated to objective cognitive performance among nongeriatric adults presenting for outpatient neuropsychological evaluation. *The Clinical*

Neuropsychologist, 1–24. https://doi.org/10.1080/13854046.2023.2241190

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

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*(3), 218–224. https://psycnet.apa.org/doi/10.1037/1040-3590.6.3.218

Harp, J. P., Jasinski, L. J., Shandera-Ochsner, A. L., Mason, L. H., & Berry, D. T. (2011). Detection of malingered ADHD using the MMPI-2-RF. *Psychological Injury and Law*, *4*(1), 32–43. https://doi.org/10.1007/s12207-011-9100-9

Harrison, A. G., & Armstrong, I. T. (2016). Development of a symptom validity index to assist in identifying ADHD symptom exaggeration or feigning. *The Clinical Neuropsychologist*, *30*(2), 265–283. https://doi.org/10.1080/13854046.2016.1154188

Harrison, A. G., Edwards, M. J., & Parker, K. C. (2007). Identifying students faking ADHD: Preliminary findings and strategies for detection. *Archives of Clinical Neuropsychology*, *22*(5), 577–588. https://doi.org/10.1016/j.acn.2007.03.008

Hirsch, O., Fuermaier, A. B., Tucha, O., Albrecht, B., Chavanon, M. L., & Christiansen, H. (2022). Symptom and performance validity in samples of adults at clinical evaluation of ADHD: A replication study using machine learning algorithms. *Journal of Clinical and Experimental Neuropsychology*, *44*(3), 171–184. https://doi.org/10.1080/13803395.2022.2105821

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

Jasinski, L. J., Harp, J. P., Berry, D. T., Shandera-Ochsner, A. L., Mason, L. H., & Ranseen, J. D. (2011). Using symptom validity tests to detect malingered ADHD in college students. *The Clinical Neuropsychologist*, *25*(8), 1415–1428. https://doi.org/10.1080/13854046.2011.630024

Lakhan, S. E., & Kirchgessner, A. (2012). Prescription stimulants in individuals with and without attention deficit hyperactivity disorder: Misuse, cognitive impact, and adverse effects. *Brain and Behavior*, *2*(5), 661–677. https://doi.org/10.1002/brb3.78

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*, *29*(6), 1522–1529. 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*(7), 1204–1237. https://doi.org/10.1080/13854046.2010.514290

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

Mitchell, J., & Read, J. (2012). Attention-deficit hyperactivity disorder, drug companies and the internet. *Clinical Child Psychology and Psychiatry*, *17*(1), 121–139. https://doi.org/10.1177/1359104510396432

Morris, N. M., Lee, T. T., Demakis, G. J., & Park, S. (2022). Detecting feigned ADHD in college students using the Minnesota multiphasic personality inventory-2-restructured form (MMPI-2-RF). *The Clinical Neuropsychologist*, *37*(6), 1–19. https://doi.org/10.1080/13854046.2022.2112294

Nitch, S. R., & Glassmire, D. M. (2007). Non-forced choice measures to detect noncredible cognitive performance. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 29–49). Guilford Publications.

Ovsiew, G. P., Cerny, B. M., Boer, A. B., Petry, L. G., Resch, Z. J., Durkin, N. M., & Soble, J. R. (2023). Performance and symptom validity assessment in attention deficit/hyperactivity disorder: Base rates of invalidity, concordance, and relative impact on cognitive performance. *The Clinical Neuropsychologist*, *37*(7), 1498–1515. https://doi.org/10.1080/13854046.2022.2162440

Phillips, M. S., Wisinger, A. M., Lapitan, F., Ausloos-Lozano, J. E., Bing-Canar, H., Durkin, N. M., Ovsiew, G. P., Resch, Z. J., Jennette, K. J., & Soble, J. R. (2023). Cross-validation of multiple embedded validity indices in the Rey Auditory Verbal Learning Test and Brief Visuospatial Memory Test-Revised in an adult attention-deficit/hyperactivity disorder clinical sample. *Psychological Injury and Law*, *16*(1), 27–35. https://doi.org/10.1007/s12207-022-09443-3

Power, M., Fell, G., & Wright, M. (2013). Principles for high-quality, high-value testing. *BMJ Evidence-Based Medicine*, *18*(1), 5–10. https://doi.org/10.1136/eb-2012-100645

Pringsheim, T., Ganos, C., McGuire, J. F., Hedderly, T., Woods, D., Gilbert, D. L.,Piacentini, J., Dale, R. C., &Martino, D. (2021). Rapid onset functional tic-like behaviors in young females during the COVID-19 pandemic. *Movement Disorders*, *36*(12), 2707–2713. https://doi.org/10.1002/mds.28778

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-stage 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

Robinson, E. V., & Rogers, R. (2018). Detection of feigned ADHD across two domains: The MMPI-2-RF and CAARS for faked symptoms and TOVA for simulated attention deficits. *Journal of Psychopathology and Behavioral Assessment*, *40*(3), 376–385. https://doi.org/10.1007/s10862-017-9640-8

Rogers, R. (1997). Researching dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 398–426). Guilford.

Rosenfeld, B., Sands, S. A., & Gorp, W. G. V. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical*

Neuropsychology, *15*(4), 349–359. https://doi.org/10.1093/arclin/15.4.349

Sagar, S., Miller, C. J., & Erdodi, L. A. (2017). Detecting feigned attention-deficit/hyperactivity disorder (ADHD): Current methods and future directions. *Psychological Injury and Law*, *10*(2), 105–113. https://doi.org/10.1007/s12207-017-9286-6

Schroeder, R. W., Martin, P. K., & Walling, A. (2019). Neuropsychological evaluations in adults. *American Family Physician*, *99*(2), 101–108.

Sollman, M. J., Ranseen, J. D., & Berry, D. T. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, *22*(2), 325. https://doi.org/10.1037/a0018857

Suhr, J. A., Buelow, M., & Riddle, T. (2011). Development of an Infrequency Index for the CAARS. *Journal of Psychoeducational Assessment*, *29*(2), 160–170. https://doi.org/10.1177/0734282910380190

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*(3), 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., & 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*, 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

Tse, P. K. Y., Finley, J.-C. A., Frick, L., Guilfoyle, J., Brooks, J., Khalid, E., Charara, R., Resch, Z., Ulrich, D., Ovsiew, G. P., & Soble, J. (2023). Cross-validating the embedded performance validity indicators in the Rey Auditory Verbal Learning Test in mixed neuropsychiatric and attention deficit/hyperactivity disorder clinical samples. *Psychology & Neuroscience*, *16*(2), 125–137. https://doi.org/10.1037/pne0000302

Tucha, L., Fuermaier, A. B. M., Koerts, J., Groen, Y., & Thome, J. (2015). Detection of feigned attention deficit hyperactivity disorder. *Journal of Neural Transmission*, *122*(S1), 123–134. https://doi.org/10.1007/s00702-014-1274-3

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

Yeung, A., Ng, E., & Abi-Jaoude, E. (2022). TikTok and attention-deficit/hyperactivity disorder: A cross-sectional study of social media content quality. *The Canadian Journal of Psychiatry*, *67*(12), 899–906. https://doi.org/10.1177/07067437221082854