Joint consideration of validity indicators embedded in Conners' Adult ADHD Rating Scales (CAARS)

Becke, Miriam; Tucha, Lara; Weisbrod, Matthias; Aschenbrenner, Steffen; Tucha, Oliver; Fuermaier, Anselm

Joint Consideration of Validity Indicators Embedded in Conners' Adult ADHD Rating Scales (CAARS)

Miriam Becke[1] · Lara Tucha[1,2] · Matthias Weisbrod[3,4] · Steffen Aschenbrenner[5] · Oliver Tucha[1,2,6] · Anselm B. M. Fuermaier[1]

Abstract

A decade of research has both illustrated the need for accurate clinical assessment of adult ADHD and brought forward a series of validity indicators assisting this diagnostic process. Several of these indicators have been embedded into Conners' Adult ADHD Rating Scales (CAARS). As their different theoretical underpinnings offer the opportunity of possible synergy effects, the present study sought to examine whether the item- or index-wise combination of multiple validity indicators benefits classification accuracy. A sample of controls ($n = 856$) and adults with ADHD ($n = 72$) answered the CAARS, including the ADHD Credibility Index (ACI) honestly, while a group of instructed simulators ($n = 135$) completed the instrument as though they had ADHD. First, original CAARS items, which are part of the CAARS Infrequency Index (CII), and items drawn from the ACI were combined into a new CII-ACI-Compound Index. Secondly, existing validity indicators, including suspect T-score elevations and the CII, were considered in combination. Both approaches were evaluated in terms of sensitivity and specificity. The combination of four CII and five ACI items into the CII-ACI-Compound Index yielded a sensitivity between 41 and 51% and an estimated specificity above 87%. Suspect T-score elevations on all three DSM scales emerged as another potentially useful validity indicator with a sensitivity of 45 to 46% and a specificity > 90%. Deeming examinees non-credible whenever two or more validity indicators showed suspect results ensured low false-positive rates (< 10%), but reduced sensitivity significantly. Classifying respondents as non-credible as soon as any given indicator fell into the suspect range resulted in frequent false positives (> 11% of misclassified adults with ADHD). Depending on whether high specificity or high sensitivity is prioritized, such combined considerations offer valuable additions to individual validity indicators. High sensitivity provided by "either/ or" combinations could prove useful in screening settings, whereas high stakes settings could benefit from "and" combinations.

Keywords Attention deficit hyperactivity disorder · Conners Adult ADHD Rating Scales · Feigning · Non-credible symptom report · Symptom validity

There is a high need for efficient validity indicators in the clinical assessment of adult ADHD, as an ever-growing evidence base suggests that individuals motivated to feign the disorder are able to do so convincingly on a variety of instruments routinely used in clinical practice. Whether they are motivated by the prospect of being prescribed stimulant medication,

✉ Miriam Becke
  m.becke@rug.nl

1 Department of Clinical and Developmental Neuropsychology, Faculty of Behavioural and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

2 Department of Psychiatry and Psychotherapy, University Medical Center Rostock, Gehlsheimer Str. 20, 18147 Rostock, Germany

3 Department of Psychiatry and Psychotherapy, SRH Clinic Karlsbad-Langensteinbach, 76307 Karlsbad, Germany

4 Department of General Psychiatry, Center of Psychosocial Medicine, University of Heidelberg, 69115 Heidelberg, Germany

5 Department of Clinical Psychology and Neuropsychology, SRH Clinic Karlsbad-Langensteinbach, 76307 Karlsbad, Germany

6 Department of Psychology, Maynooth University, National University of Ireland, Maynooth, Ireland

receiving accommodations at school or work, or benefitting in social interactions (Fuermaier et al., 2021), adults feigning ADHD commonly score in the clinically relevant yet believable range on self-report instruments (Fuermaier et al., 2016a, b; Fuermaier et al., 2017a, b; Harrison & Armstrong, 2016; Harrison et al., 2007; Jachimowicz & Geiselman, 2004; Booksh et al., 2010; Marshall et al., 2016; Quinn, 2003; Smith et al., 2017; Walls et al., 2017). A recent survey among Dutch university students found 57% of them to consider feigning ADHD an easy endeavor, and nearly 52% of them expected an illegitimate diagnosis of ADHD to bring about benefits (Fuermaier et al., 2021). Respondents most commonly believed individuals with such a diagnosis to benefit from access to accommodations in the academic context, followed by prescription of stimulant medication, or advantages in social interactions (e.g., receiving attention, having an excuse for occasional misconduct). These numbers appear to translate into high base rates of feigned ADHD, particularly among university students. In an early study, Harrison and colleagues (2007) found 8% of their ADHD sample to be classified as faking and suggested this to correspond to the base rate of feigned ADHD in their particular sample of undergraduate students. This estimate has since then been revised upwards, and some expect it to rise further in the future (Sagar et al., 2017). To the best of our knowledge, most recent studies have noted base rates converging around 20%, though some describe rates as high as 50% (Marshall et al., 2010; Martin & Schroeder, 2020; Suhr et al., 2008; Sullivan et al., 2007).

Studies examining the use of self-report questionnaires and clinical interviews in the detection of feigned ADHD have been cornerstones to efforts aimed at identifying feigned instances of the disorder. Even though such instruments are central to the diagnostic process, owing to the subjective nature of the symptoms characterizing ADHD, few include validity indicators. Conner's Adult ADHD Rating Scales (CAARS) (Conners et al., 1999) are among the few disorder-specific instruments with embedded validity indicators. One such indicator, the CAARS' Inconsistency Index, aims to uncover careless or random responding. As underscored by Harrison et al. (2007), this index is not intended to identify individuals who "fabricate symptoms for purposes of secondary gain."

Evidence suggesting that individuals feigning ADHD tend to over-endorse symptoms (i.e., present with higher scores or levels of symptomatology than credible patients with ADHD; Harrison et al., 2007; Suhr et al., 2011) illustrates the possible use of another validity indicator described in the CAARS manual (Conners et al., 1999). The authors state that T-scores exceeding 80 should be considered suspect. Although they may arise from particularly severe symptomatology, such elevated scores warrant further investigation of possible symptom exaggeration or malingering.

In 2011, Suhr and colleagues introduced the first embedded validity indicator aiding the detection of feigned ADHD, which is based on infrequent as opposed to inconsistent symptom reports. Their CAARS Infrequency Index (CII) includes original CAARS items which are rarely endorsed by genuine patients with ADHD and neurotypical controls (see "Method" for additional details). As such, it is particularly suitable for the detection of non-credible symptom overreport. Cross-validations of the CII revealed variable, but promising classification accuracy (Cook et al., 2016, 2017; Edmundson et al., 2017; Fuermaier et al., 2016a, b; Harrison & Armstrong, 2016; Walls et al., 2017).

Five years later, Harrison and colleagues reported results on a new index, the Exaggeration Index (EI), which embedded items from the Dissociative Experiences Scale (DES) (Bernstein & Putnam, 1986) into the CAARS. Early data on the EI's classification accuracy were encouraging; cross-validations are pending though.

More recent developments include the Dissimulation ADHD Scale (Ds-ADHD) introduced by Robinson and Rogers (2018), the ADHD Symptom Infrequency Scale (ASIS) with its Infrequency Scale (INF) (Courrégé et al., 2019), and the ADHD Credibility Index (ACI) (Becke et al., 2021). They share a common conceptual basis in that all three scales tap misconceptions about ADHD. The Ds-ADHD draws these stereotypes from the MMPI-2-RF (Ben-Porath & Tellegen, 2008) by including only those items believed to be indicative of ADHD by people without a secured diagnosis of the disorder, but not credible adults with an established diagnosis of ADHD. While the CII, EI, and Ds-ADHD utilize items drawn from existing instruments, the INF and the ACI each contain disorder-specific items written specifically for the purpose of detecting non-credible self-report in the clinical assessment of ADHD. The ACI's items were developed as ADHD-specific adaptations of detection strategies described by Rogers (Becke et al., 2019; Rogers, 2018). Initial validation studies showed comparable classification accuracy for the Ds-ADHD and INF, whereas the performance of the ACI was comparable to the CII's classification accuracy.

Despite comparable sensitivity, early data on the ACI indicate that it may identify a different subset of feigning individuals than the CII (Becke et al., 2021). Given that different validity indicators detect different cases of non-credible self-report, considering multiple indicators jointly could increase sensitivity to instances of feigned ADHD. The CII and ACI lend themselves to such a combined consideration as their development was based on different approaches and theoretical underpinnings. Their items thus cover distinct content, suggesting their integration may benefit the detection of non-credible self-report.

The present study aimed to determine whether various validity indicators embedded in the CAARS could be combined in a manner that increases sensitivity while continuing to ensure high specificity. We investigated this possibility using a simulation design: a group of adults with ADHD, a sample of adults without a secured diagnosis of ADHD, and a group of instructed simulators completed the CAARS including the ACI.

Two approaches were followed to find a suitable combination of suspect elevations on the CAARS DSM scales (see "Method" for further information), the CII, and the ACI. First, we formed a CII-ACI-Compound Index including all items which maximize differences between instructed simulators and genuine cases of adult ADHD and determined a suitable cut-off score for this CII-ACI-Compound Index. Second, we analyzed configural features which discriminate simulators from credible adults with ADHD, such as *T*-scores exceeding 80 on at least one of the CAARS' three DSM scales in conjunction with suspect scores on the CII or ACI.

Method

The study design and data set described here were adapted from Becke et al. (2021) to investigate the previously unexamined, combined consideration of multiple validity indicators. For this purpose, participant selection and methods were modified as outlined in the following sections.

Participants

Control Group

The *control group* included 1577 panel members registered with a Dutch online platform, which offers financial compensation for participation in web-based research studies. Approximately 30% of this initial sample ($n = 460$) withdrew from participation before having answered all questionnaires, and an additional 7.42% of respondents were excluded due implausibly short time needed to complete the instruments under investigation ($n = 117$). Thirty-five CAARS protocols (2.22%) were dismissed due to five or more missing answers, and 46 participants in the *control group* (2.92%) responded inconsistently (i.e., CAARS Inconsistency Index ≥ 8) and were therefore excluded from further analyses. Respondents were included such that the sample presented a cross-section of the Dutch population in terms of age, sex, and education (see Table 1). However, 18 participants in the *control group* (1.14%) reported a neurological or psychiatric morbidity warranting exclusion, and 45 respondents (2.85%) were excluded due to the recent intake of medications known to affect the central nervous system.

We divided the remainder of this group into a larger subsample, which encompassed 75% of the remaining 856 participants ($n = 642$), and a smaller subsample, which included 25% of controls ($n = 214$). The larger subsample, henceforth called C-1, served in the development and initial validation of both compound index and configural features. The smaller subset, from here on termed C-2, was analyzed as part of cross-validation.

Participants allocated to subset C-2 did not differ from the controls in subset C-1 in terms of age ($z = -1.203$, adjusted $p = 1.00$), gender distribution ($\chi^2 (1) = 0.000$, $p = 1.00$), or education ($z = 0.017$, adjusted $p = 1.00$). Demographic data of both subsets are shown in Table 1.

ADHD Group

The *ADHD group* initially considered 122 adults with ADHD, who had been referred to the Department of Psychiatry and Psychotherapy at the SHR Clinic in Karlsbad-Langensteinbach, Germany, by local psychiatrists or neurologists. This specialized outpatient clinic offers thorough diagnostic workup to adults whose general practitioners, psychiatrists, or neurologists suspect the presence of ADHD but do not feel sufficiently experienced or qualified to diagnose the disorder in adulthood. Even though all examinees had experienced symptoms and impairments of ADHD throughout childhood and adolescence, it could not be conclusively determined whether a formal diagnosis has been established in childhood or not for most of the cases. The diagnostic procedure, therefore, followed criteria for first-time adult ADHD diagnosis (e.g., see Sibley, 2021). As outlined in our previous study (Becke et al., 2021), where this sample was first described, a comprehensive clinical work-up conducted by at least two experienced clinicians secured their diagnoses. It included a psychiatric interview which enquired both past and present symptoms and impairments of ADHD as laid out in the DSM (American Psychiatric Association, 2000, 2013; Barkley & Murphy, 1998), as well as the completion of two standardized self-report rating scales which tapped symptoms of ADHD across the same period (WURS-K and ASR) (Adler et al., 2006; Kessler et al., 2005; Ward et al., 1993). Corroborating evidence of impairments in line with ADHD was also collected by asking parents, partners, and/or employers about difficulties apparent at school, home, or work. Objective evidence of impairment was further indicated by academic failure, negative teacher evaluations, unstable employment, financial problems, frequent relationship break-ups, repeated legal incidences, and poor driving records. Retrospective accounts of symptoms and impairments in line with ADHD experienced during childhood and adolescence are a prerequisite for a first-time diagnosis of the disorder in adulthood, such that no formal diagnosis was given without such evidence.

Table 1 Descriptive data by group

	Control group (n = 856)		ADHD Group (n = 72)		Simulation group (n = 135)	
	Subset C-1 (n = 642)	Subset C-2 (n = 214)	Subset A-1 (n = 54)	Subset A-2 (n = 18)	Subset S-1 (n = 102)	Subset S-2 (n = 33)
Age (years)						
Median (MAD)	49 (11)	51 (10)	34 (10)	36 (7)	20 (1)	19 (1)
Range (Min–Max)	40 (25–65)	40 (25–65)	62 (18–80)	35 (19–54)	40 (18–58)	38 (17–55)
Sex (m/f)	319/322*	106/107*	27/27	5/13	22/80	9/24
%	49.69/50.16	49.53/50.00	50.00/50.00	27.78/72.22	21.57/78.43	27.27/72.73
Education						
Years						
Median (MAD)	10 (1)	10 (1)	12 (3)	14 (4)	14 (1)	14 (1)
Range (Min–Max)	10 (9–19)	10 (9–19)	15 (9–24)	13 (8–21)	14 (11–25)	13 (12–25)
ADHD symptomatology						
Past[a]						
Median (MAD)			42 (10)	37 (9)	14 (7)	12 (6)
Range (Min–Max)			60 (10–70)	66 (0–66)	39 (0–39)	48 (0–48)
Present[b]						
Median (MAD)			33 (8)	35 (6)	10 (4)	11 (6)
Range			53 (0–53)	45 (0–45)	41 (0–41)	45 (0–45)

MAD median absolute deviation

[a]Wender Utah Rating Scale

[b]ADHD Self-Report Scale

*two participants did not specify their gender

All participants in this group were also administered at least one independent performance validity test (Test of Memory Malingering [TOMM] introduced by Tombaugh, 1996; Groningen Effort Test [GET] developed by Fuermaier et al., 2017b; Fuermaier et al., 2016a, b). Twenty-two participants (18.03%) were excluded from the group due to suspect results (i.e., five of 59 individuals failing the TOMM, and 17 of 63 individuals failing the GET). Twenty-eight adults with ADHD (22.95%) answered the CAARS in an inconsistent manner (i.e., CAARS Inconsistency Index ≥ 8) and were consequently excluded, resulting in a final sample size of 72 adults with ADHD.

Participants in the *ADHD group* most commonly met the diagnostic criteria for the combined ADHD subtype (n = 38, 52.78%), followed by the inattentive type (n = 28, 38.89%). The subtype remained unspecified for six cases (8.33%). Psychiatric and neurological comorbidities were common among adults with ADHD, which is typical for this population and referral context (Biederman et al., 1993; Chen et al., 2018; Cumyn et al., 2009), especially if ADHD was unrecognized or insufficiently treated earlier in life. In the present sample, 36% of participants reported at least one such comorbid condition, while 26% of adults with ADHD reported two or more comorbidities. Mood (73% of all diagnoses) and anxiety disorders (22% of all diagnoses)

accounted for most of these comorbidities. Personality disorders (7% of all diagnoses) and substance use disorders (4% of all diagnoses) were less common.

Paralleling our approach described for the *control group*, 25% of group members were randomly assigned to a subset of participants whose data were to be used for cross-validation. This smaller subgroup (n = 18) was again denoted A-2, whereas we will refer to the larger subsample of the *ADHD group* (n = 54) as A-1. As was the case for the *control group*, subset A-2 of adults with ADHD did not differ from subset A-1 in age (z = 0.345, adjusted p = 1.00), gender distribution (χ^2 (1) = 2.700, p = 0.100), or education (z = −1.391, adjusted p = 1.00). Participants in the combined *ADHD group* did, however, differ significantly from those in the *control group* with regard to age (z = −6.086, adjusted p < 0.01). There were no significant group differences in gender distribution (χ^2 (1) = 0.752, p = 0.386) or years of education (z = 1.199, adjusted p = 0.692). Demographic data are summarized in Table 1.

Simulation Group

The *simulation group* was recruited through public announcements and researchers' contacts, as well as word-of-mouth, and included 174 adults asked to feign ADHD. Participants

were *not* excluded if they evidenced symptoms of ADHD; however, five examinees in this group (2.87%) reported a history of psychiatric or neurological conditions other than ADHD and were consequently excluded from further analyses. Thirty-four instructed simulators (19.54%) were excluded as they had answered the CAARS in an inconsistent manner (i.e., CAARS Inconsistency Index ≥ 8).

Twenty-five percent of the 135 remaining instructed simulators ($n=33$) were chosen at random and allocated to subset S-2 (see Table 1). This subset of simulating participants did not differ from subset S-1 ($n=102$) of the simulating sample in age ($z=0.379$, adjusted $p=1.00$), gender distribution (χ^2 (1) = 0.459, $p=0.498$), or education ($z=0.074$, adjusted $p=1.00$).

The combined *simulation group* was, on average, significantly younger than participants in the *ADHD group* ($z=6.394$, adjusted $p<0.01$) and the *control group* ($z=18.138$, adjusted $p<0.01$). Gender distribution within the *simulation group* differed significantly from both the *ADHD group* (χ^2 (1) = 10.234, $p<0.01$) and the *control group* (χ^2 (1) = 33.702, $p<0.01$). Overall, instructed simulators had completed more years of education than controls ($z=-7.370$, adjusted $p<0.01$) and adults with ADHD ($z=-3.598$, adjusted $p<0.01$). Table 1 summarizes all demographic data.

Materials

ADHD Symptom Severity

Participants indicated the presence and severity of ADHD symptomatology during childhood on the Wender Utah Rating Scale (WURS-K) (Ward et al., 1993). Its short-form includes 25 items, which assess symptoms of ADHD experienced between the ages of 8 and 10 years. Items are rated on a 5-point scale, ranging from 0 ("Does not apply") to 4 ("Strong manifestation"). All items except numbers 4, 12, 14, and 25 are summed up to form a total score, which is considered to indicate clinically significant symptomatology if it exceeds the recommended cut-off score of 30.

Current symptoms of ADHD were measured using the ADHD self-report scale (ASR) (Adler et al., 2006; Kessler et al., 2005). The ASR consists of 18 items assessing symptoms of ADHD as laid out in the DSM-IV. Participants indicate their answer on a 4-point scale ranging from 0 ("Does not apply") to 3 ("Strong manifestation"). The sum of all items suggests clinically relevant symptoms of ADHD if it surpasses 18.

Experimental Version of Conners' Adult ADHD Rating Scale (CAARS-ACI)

The long-form of Conners' Adult ADHD Rating Scales (CAARS) (Conners et al., 1999) contain 66 items which assess the presence and severity of ADHD symptomatology by presenting participants with statements pertaining to everyday activities and tendencies in behavior. Examinees indicate the extent to which these statements apply on a 4-point scale ranging from 0 ("not at all/never") to 3 ("very much/very frequently"). The CAARS includes factor-derived scales, which tap inattention and memory problems, hyperactivity and restlessness, and impulsivity and emotional lability, as well as participants' self-concept. ADHD symptoms as listed in the DSM-IV (American Psychiatric Association, 2000) are measured by three additional scales and summarized in the CAARS DSM Total. The CAARS also allows for the calculation of an ADHD Index with a reported sensitivity of 87% and a specificity of 85% in the detection of clinically relevant symptoms of ADHD (Conners et al., 1999). Lastly, the CAARS includes an Inconsistency Index, which suggests careless or random responding if it exceeds a score of 7.

In addition to inconsistent responding, endorsement of infrequent symptoms may be gauged using the CAARS and used to aid the detection of non-credible self-report. The CAARS Infrequency Index (CII), introduced by Suhr et al. in 2011, comprises CAARS items that are infrequently endorsed by healthy controls and credible adults with ADHD. Early studies of the CII showed a sensitivity of 24% and a specificity of 95% to instances of feigned vs. genuine adult ADHD. In subsequent cross-validations, the sensitivity of the CII ranged from 17% or 18% (Cook et al., 2016, 2017), 34% (Walls et al., 2017) to approximately 50% (Fuermaier et al., 2016a, b; Robinson & Rogers, 2018). Specificity of the CII has been found to be high, ranging from 86% (Robinson & Rogers, 2018) to 95% (Walls et al., 2017; see Fuermaier et al., 2016a, b for an exception).

As part of an earlier study, we developed the ADHD Credibility Index (ACI) as a disorder-specific validity indicator embedded in the CAARS and subjected it to an initial validation (Becke et al., 2021). The final index includes 12 items, all of which are based on previously described detection strategies (see Rogers, 2008) and can be subsumed in four subscales: selective symptoms, supposed symptoms, exaggerated symptoms, and symptom combinations. Applying a conservative cut-off score of 21, the ACI sum score distinguished instructed simulators from credible cases of adult ADHD with a sensitivity of 44% and a specificity of 98%.

Tests of Performance Validity in ADHD Groups

All adults with ADHD completed at least one performance validity test as part of the diagnostic process. Results of the Test of Memory Malingering (TOMM) (Tombaugh, 1996) or the Groningen Effort Test (GET) (Fuermaier et al., 2016a,

b, 2017a, b) were considered to ensure that the *ADHD group* would include only those examinees whose data provided no evidence of non-credibility.

Results yielded by the TOMM, a visual memory recognition test that uses a forced-choice format and floor effects to detect non-credible performance, were considered suspect if participants correctly identified fewer than 45 of 50 items on trials 1 or 2. Applying this cut-off score, the TOMM's sensitivity is 56% and its specificity 93% (Greve et al., 2006).

The GET, on the other hand, is a computerized visual discrimination task designed to appear cognitively taxing. While examinees are suggested the task places high demands on attention and concentration, it is actually completed with ease by most individuals – including those with ADHD. Errors and reaction times are combined into an overall score, which is subsequently classified as suspect if it exceeds the cut-off value. The GET's sensitivity and specificity have been reported at 89% (Fuermaier et al., 2017a, b).

Procedure

Neurotypical Control Group

Participants in the *neurotypical control group* gave written informed consent before providing demographic information (i.e., age, sex, and educational attainment) and relevant medical history (i.e., history of psychiatric or neurological disease; pharmacological treatments affecting the central nervous system). They were subsequently instructed to complete all self-report measures (i.e., WURS-K, ASR, CAARS-ACI) honestly and to the best of their ability. The assessment procedure for healthy participants was approved by the Ethical Committee Psychology (ECP) at the University of Groningen.

ADHD Group

Adults with ADHD gave written informed consent and provided anamnestic as well as health history data before being tested individually in a quiet room on clinic premises. They were informed that all data collected as part of the study would be analyzed anonymously and that the result would not affect their clinical assessment or treatment. Adults with ADHD received no compensation for participation in the research project.

Patients underwent a comprehensive clinical assessment, which included routine measures of cognition alongside the self-report questionnaires and validity tests described previously. Testing took approximately 2 h, divided into two parts to avoid the potential effects of fatigue (Lezak et al., 2004). The study complied with the ethical standards of the Helsinki Declaration and was approved by the local institutional ethical committee (Medical Faculty at the University of Heidelberg, Germany).

Simulation Group

Participants in the *simulation group* gave written informed consent and provided the same demographic information and medical history as honest-responding controls. In contrast to the *control group*, however, they were instructed to answer anything but these inquiries into their demographic and medical backgrounds as though they had ADHD. That is, the WURS-K, ASR, and CAARS-ACI were completed while instructed to feign ADHD. As an independent validity test, they additionally took the TOMM or GET. The instruction to feign ADHD believably (e.g., by avoiding pronounced exaggeration of symptoms) was incentivized through the chance of winning a tablet PC: the participant who simulated ADHD most convincingly was to be awarded this PC. Unbeknownst to participants, the winner was chosen at random, i.e., independent of test results. Verbatim instructions given to participants in the *simulation group* are available from the corresponding author upon request.

To help them adopt the role of an adult with ADHD, volunteers received vignettes ahead of the assessment, which described multiple possible incentives for someone to simulate the disorder (e.g., financial, educational, or vocational accommodations, or the prescription of stimulant medication). Examiners were aware of the instructions the simulating participants had received. Participants took approximately 70 min to answer questions about their demographic and medical backgrounds and complete all self-report measures as well as performance validity testing. After completion, participants were debriefed and instructed to stop feigning ADHD. They were asked whether they had followed the given instructions. All participants answered in the affirmative.

Statistical Analyses

The following analyses were first conducted using data provided by subsets 1 of the *control*, *ADHD*, and *simulation groups* (see Table 1). Subsets 2 were *not* considered in the development and initial analyses of the CII-ACI-Compound Index and configural features but their subsequent cross-validation.

Compound Index We first examined the possibility of synergetic effects between different validity indicators on the level of individual items included in the CII and ACI. In its initial validation, the ACI showed a sensitivity of approximately 44% to instances of feigned ADHD and a large effect (Sullivan & Feinn, 2012) for the comparison of credible adults with ADHD and instructed simulators (Becke et al., 2021). We raised each individual item to the standard fulfilled by the ACI sum score – that is, a large effect size – and combined items that met this standard into a compound index. That is, items were selected from the pool of

all available CII and ACI items, if (a) Cohen's *d* was equal to or higher than 0.80 and (b) simulators tended to produce higher scores on these items than adults with ADHD.

Answers to the selected items were summed up using the CAARS' original 4-point scale. We detailed descriptive statistics on the compound index's sum scores. Possible group differences in these scores were examined using nonparametric significance tests, and we report Cohen's *d* for the comparison of adults with ADHD and instructed simulators. We further determined a cut-off score that would ensure at least 90% specificity. Sensitivity and specificity as well as positive (PPV) and negative predictive value (NPV) estimates of the new compound index were calculated as preliminary data on its classification accuracy. ROC analyses were run to gauge the compound index's accuracy in detecting instructed simulators as well as overreporting participants.

Configural Features Secondly, we considered the classification accuracies of validity indicators embedded in the CAARS and their joint consideration. Data inspection revealed what Harrison and Armstrong (2016) termed "configural features": results or data patterns which occur more frequently among simulators than credible cases of adult ADHD. We determined the percentages of instructed simulators and adults with ADHD identified by each of these indicators and investigated whether a combination of them would allow for higher sensitivity rates than individual ones while also ensuring high specificity. Validity indicators under investigation included suspect *T*-score elevations on the DSM scales, the CII, and the ACI. To allow for these analyses, participants remained in the sample if they presented with suspect scores on any given one of them.

Results

Compound Index

As illustrated by Fig. 1, effect sizes observed for the comparison of instructed simulators and credible adults with ADHD varied from negligible to large. Eight items met the criterion of a large effect for the comparison of adults with ADHD and instructed simulators (i.e., Cohen's $d \geq 0.80$). Four original CAARS items, which are part of the CII (i.e., item numbers 21, 22, 45, and 62), and five ACI items showed the largest effects. The single largest effect was observed for ACI item number 10 (corresponding to item number 58 in Becke et al., 2021). Approximately 13% of all adults with ADHD ($n = 9$) endorsed it as occurring "pretty much, often" to "very much, very frequently" (12.96%, $n = 7$ in A-1; 11.76%, $n = 2$ in A-2), whereas it detected approximately 70% of all instructed simulators ($n = 94$; 66.67%, $n = 68$ in S-1; 78.79%, $n = 26$ in subset S-2).

The internal consistency of the resulting index was high (Cronbach's $\alpha = 0.95$). With each item being scored on a scale from 0 to 3, possible sum scores on the CII-ACI-Compound Index ranged from 0 to 27. Table 2 shows descriptive statistics on the new index, with significantly higher average sum scores for members of the *simulation group* compared to the *ADHD* (z = − 4.329, adjusted $p < 0.01$) or *control groups* (z = − 16.264, adjusted $p < 0.01$). The effect size resulting from the comparison of simulators and credible adults with ADHD was very large by Rogers' standards (Rogers, 2018) and large by Cohen's classification (Sullivan & Feinn, 2012) ($d = 1.551$; 95%-CI [0.608, 2.494]).

A cut-off score of 17 was needed to ensure specificity of at least 90% among adults with ADHD, whereas a score of five sufficed to guarantee similar specificity in the *control group*. Considering scores above 17 indicative of non-credible self-report, 50.51% of simulators ($n = 50$) were marked as non-credible. In the *ADHD group*, 92.45% percent of participants ($n = 49$) fell below the cut-off score and 99.53% of controls ($n = 639$) presented with scores below the cut-off value. Sensitivity and specificity, as well as positive (PPV) and negative predictive values (NPV) of the CII-ACI-Compound Index are presented in Table 3.

Overall classification accuracy was determined in a series of ROC analyses. Using the CII-ACI-Compound Index as a criterion in differentiating credible adults with ADHD from instructed simulators resulted in an area under the curve (AUC) of 0.855 ($SE = 0.033$, $p < 0.01$, 95% CI [0.790, 0.920]; see Figs. 2 and 3). When used to identify overreport on any DSM scale (i.e., $T \geq 80$), rather than group membership, the CII-ACI-Compound Index showed an AUC of 0.785 ($SE = 0.040$, $p < 0.01$, 95% CI [0.706, 0.864]).

Configural Features

Table 4 summarizes the percentages of participants in each group identified by each validity indicator. The highest sensitivity to the instructed simulation of ADHD was observed for *T*-scores equal to or above 80 on the DSM Inattention (E) and DSM Total (G) scales alongside suspect results on the CII. These indicators identified between 67 and 74% of instructed simulators. With approximately one-third to half of all adults with ADHD being classified as non-credible, their specificity was low. *T*-score elevations on the ADHD Index (H) were marginally more common in the *ADHD group* than the *simulation group*, limiting the index's utility in the detection of instructed simulation.

High false-positive rates observed for the DSM scales and CII were driven by a subset of participants in the *ADHD group* who endorsed high levels of symptomatology (i.e., *T*-score ≥ 80 on any DSM scale). In the case of the CII, for instance, approximately 63% of overreporting participants with ADHD produced suspect scores, whereas this was
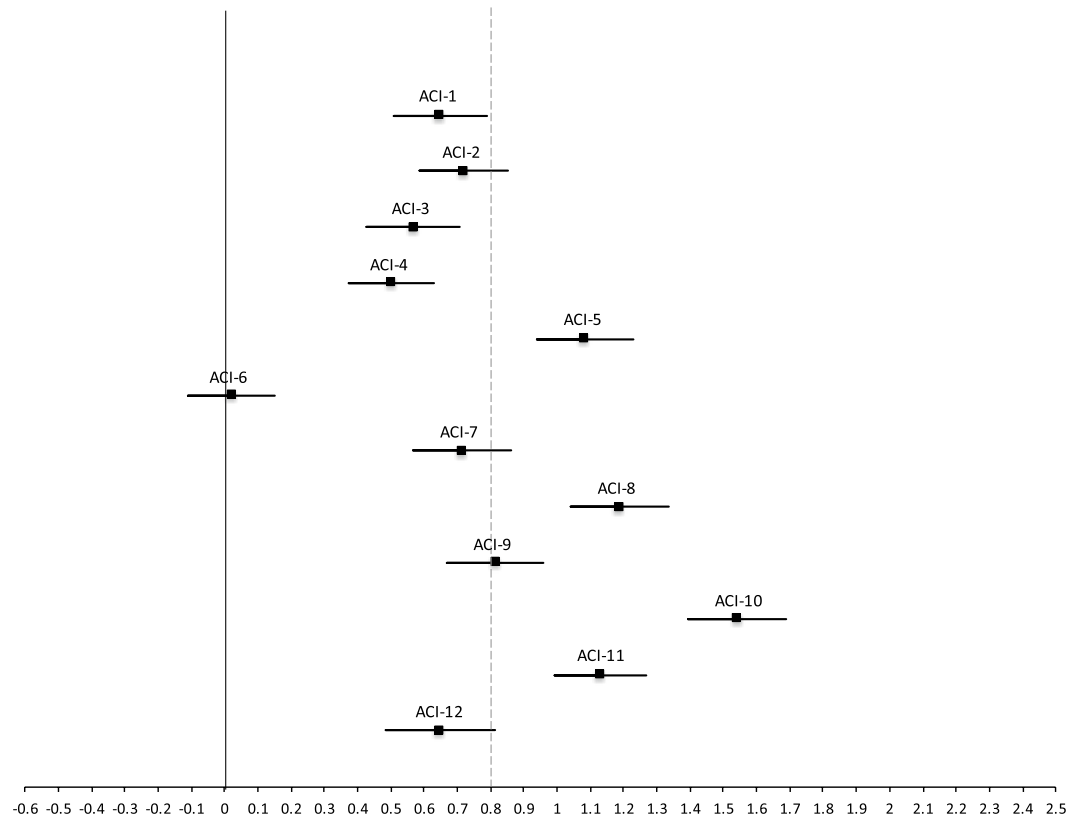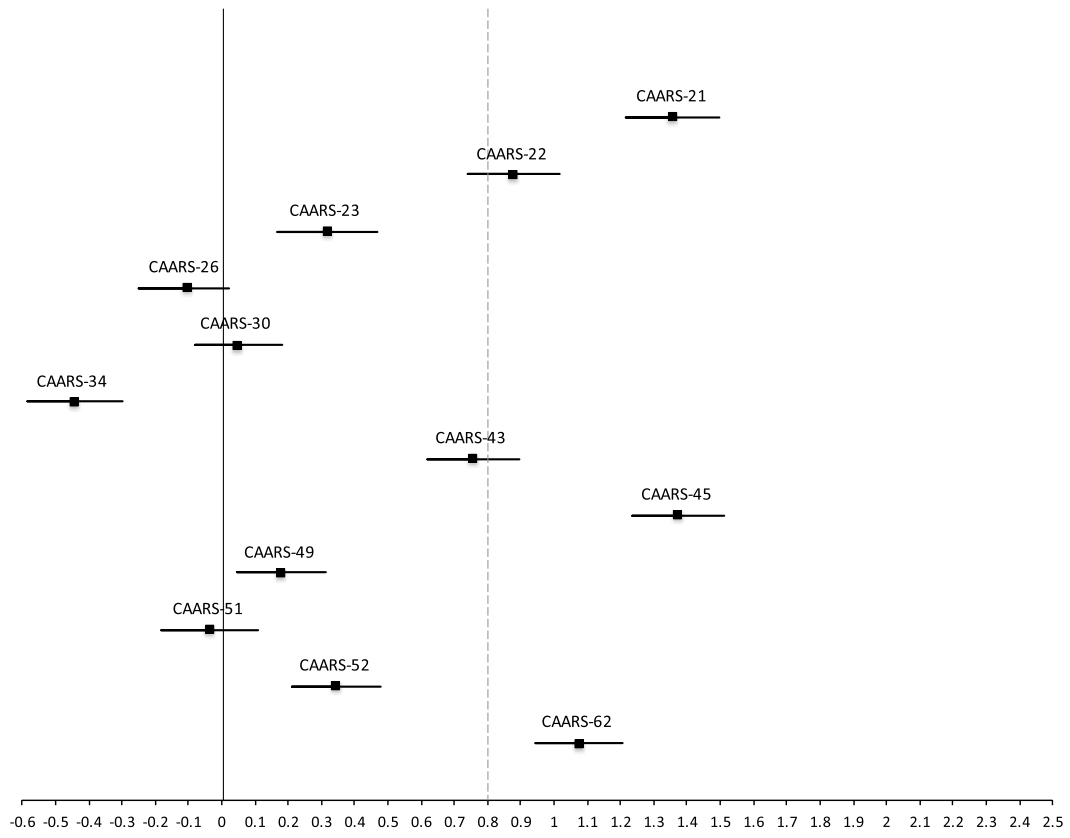
◄**Fig. 1** Forest plots illustrating effect sizes (Cohen's *d*) yielded by CII (top) and ACI (bottom) items for the comparison of the *simulation* and *ADHD group*. Bars to the right of the horizontal axis indicate higher average scores for simulators than adults with ADHD. Dashed line indicates effect size required for inclusion of items into the CII-ACI-Compound Index

the case for ca. 5% of patients with ADHD who did not overreport.

Specificity was increased when these indicators were considered in conjunction rather than isolation (see Table 5). Elevations on individual DSM scales were common among adults with ADHD. Suspect elevations on *all three* DSM scales, in contrast, occurred for ca. 7% of adults with ADHD and less than 1% of controls, while identifying 46% instructed simulators in sample S-1. The percentage of misclassified adults with ADHD was reduced by 31% when combining the CII with suspect elevations on all DSM scales (compared to the CII considered in isolation). The specificity of the ACI was increased by ca. 4% when interpreted in combination with DSM scale *T*-scores. Jointly considering the CII and ACI increased the CII's specificity by approximately 30%, corresponding to the ACI's specificity of 91%.

Requiring suspect scores on multiple validity indicators to warrant classification as non-credible decreased sensitivity. Compared to the interpretation of *T*-scores alone, combining *T*-score elevations on the DSM scales with suspect results on the CII or ACI reduced sensitivity by 3% and 15% respectively. Relative to the CII and ACI, the same combination lowered sensitivity by 24% and 19%. Joint consideration

of the CII and ACI resulted in a 23% drop in sensitivity for the CII and a 6% decrease for the ACI.

If participants were classified as non-credible whenever the CII *or* the ACI yielded suspect results, sensitivity to feigned ADHD was increased by approximately 7% compared to the isolated consideration of the CII and by approximately 24% relative to the ACI interpreted in isolation. Specificity was equal to that of CII, i.e., the least specific indicator included in this configural feature. Classifying self-report as non-credible when suspect elevations on the DSM scales co-occurred with suspect scores on either CII *or* ACI showed satisfactory specificity among adults with ADHD (ca. 92%), and a sensitivity of approximately 44%. Specificity was significantly lower than 90% for all other configural features which combined multiple validity indicators in an "either/or" fashion.

Cross-validation

Compound Index

While participants drawn from the *simulation group* for cross-validation (i.e., subset S-2) presented with higher average scores than the subset of participants sampled from the *ADHD group* (i.e., subset A-2; see Table 2), scores on the CII-ACI-Compound Index did not differ significantly between the two groups ($z = -1.496$, adjusted $p = 1.00$). Paralleling findings on subset S-1, their comparison resulted in a large effect (Cohen's $d = 1.655$; 95% CI [0.379, 2.931]) by both Rogers' or Cohen's standards (Rogers, 2018; Sullivan

Table 2 Descriptive statistics on the CII-ACI-Compound Index by group

	Median	MAD	Mode	Range	Min–Max	Mean	SD
Control group							
Total	1[b, c]	1	0	21	0–21	2.19	2.82
Subset C-1	1[b, c]	1	0	21	0–21	2.18	2.86
Subset C-2	1[b, c]	1	0	21	0–21	2.21	2.72
ADHD group							
Total	7[a, c]	4	4	25	0–25	8.09	5.88
Subset A-1	7[a, c]	4	4	25	0–25	7.83	6.04
Subset A-2	8[a]	4	6	17	2–19	8.94	5.39
Simulation group							
Total	17[a, b]	4	18[d]	24	3–27	16.86	5.54
Subset S-1	18[a, b]	5	16[d]	24	3–27	17.03	5.93
Subset S-2	17[a]	3	18	18	7–25	16.32	4.09

MAD median absolute deviation

[a]significantly different from *control group*

[b]significantly different from *ADHD group*

[c]significantly different from *simulation group*; superscripts indicate results yielded by the comparison of corresponding subsets (e.g., total *ADHD group* vs. total *simulation group*, A-1 vs. S-1, and C-2 vs. A-2, not A-1 vs. S-2). Significance values have been adjusted by the Bonferroni correction for multiple tests

[d]Multiple modes exist; the smallest value is shown

Table 3 Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the CII-ACI-Compound Index in the detection of simulated ADHD

| | Base rate | Subsamples | |
		A-1 vs. S-1	A-2 vs. S-2
Sensitivity		50.51	41.49
Specificity[a]		92.45	87.50
PPV	5	26.05	15.01
	10	42.65	27.15
	20	62.59	45.61
	30	74.15	58.98
	40	81.69	69.10
	50	87.00	77.04
NPV	5	97.26	96.63
	10	94.39	93.13
	20	88.20	85.77
	30	81.34	77.86
	40	73.70	69.33
	50	65.13	60.11

[a]Percentage of adults with ADHD who presented with unremarkable scores on the CII-ACI-Compound Index (i.e., controls not considered)

& Feinn, 2012) nonetheless. Overall classification accuracy was marginally lower in subsamples 2 compared to subsamples 1 (see Table 6), as illustrated by a decrease in AUC of 1% when aiming to detect instructed simulators among a group of adults with ADHD. This effect was largely due to a decrease in sensitivity: the new compound index detected approximately 42% of instructed simulators ($n = 13$), while 99% of controls ($n = 213$) and ca. 88% of adults with ADHD ($n = 14$) presented with unremarkable scores. A similar decrease in classification accuracy could be noted when using the CII-ACI-Compound Index to detect overreport in a pooled group of adults with ADHD and instructed simulators (see Table 6).

Configural Features

Cross-validation revealed overall lower sensitivity of most configural features in subsamples 2 relative to subsamples 1, and reduced specificity for some indicators. The CII and overreport on the DSM Inattention (E) Scale or DSM Total (G) remained the most sensitive indicators, yet they identified ca. 4 to 10% fewer instructed simulators than observed in subset 1. The ACI showed a decline of 17% in sensitivity when subsamples 2 were compared to subsets 1. Configural features, which require multiple indicators to show suspect results for an examinee to be deemed non-credible, exhibited a decrease in sensitivity of approximately 4 to 14%. To a lesser degree, such a decline could be observed for

"either/or" combinations as well. Among controls, specificity remained stable when comparing subsets 1 and subsets 2. Adults with ADHD in subset 2 were more commonly misclassified by suspect T-score elevations co-occurring on all three DSM scales than their peers in subset 1. The same held true for the CII. Overall, configural features which joined multiple indicators in an "either/or" fashion were also less specific in the classification of subsample A-2 compared to subsample A-1.

Discussion

Following a decade of research efforts, different validity indicators are now available for the CAARS. The instrument's authors initially provided the opportunity to detect unmotivated or non-compliant responding by means of the Inconsistency Index, and also proposed that T-scores exceeding 80 might implicate overreport of ADHD symptomatology (Conners et al., 1999). Later additions to these early validity indicators aimed to detect the report of infrequent or unusual symptoms (Becke et al., 2021; Courrégé et al., 2019; Harrison & Armstrong, 2016; Robinson & Rogers, 2018; Suhr et al., 2011). Consequently, all indicators available today differ in their theoretical underpinnings and preliminary evidence has implied that they may identify different subsets of examinees as non-credible (Becke et al., 2021). The present study, therefore, sought to examine whether the combined consideration of these validity indicators could improve overall classification accuracy.

Our results suggest that the detection of symptom overreport and the endorsement of implausible symptoms benefit the assessment of adult ADHD in distinct ways. Since overreport was common in our sample of instructed simulators, validity indicators that detected it were – overall – most sensitive to simulated ADHD. By contrast, the endorsement of implausible or unusual symptoms was more specific to instances of feigned ADHD. Despite their respective strengths, the item- or index-wise combination of both strategies did not increase classification accuracy significantly in the present study.

Neither the CII-ACI Compound Index (i.e., item-wise combination) nor the configural feature, which required congruent suspect results on the CII and the ACI to warrant the classification as non-credible (i.e., index-wise combination), outperformed the ACI markedly. Still ACI item number 10 (i.e., item number 58 in the CAARS-E) yielded the largest effect for the comparison of the *ADHD* and the *simulation group* in the present study, and was among the items least frequently endorsed by patients with ADHD in the ACI's initial validation (Becke et al., 2021). Endorsement of this item identified approximately 70% of instructed simulators

Fig. 2 Receiver operating characteristics (ROC) curve indicating diagnostic accuracy of the new CII-ACI-Compound Index in identifying feigned ADHD (Subset S-1 of *simulation group*, $n = 99$) relative to ADHD (subset A-1 of *ADHD group*, $n = 53$)

ROC Curve
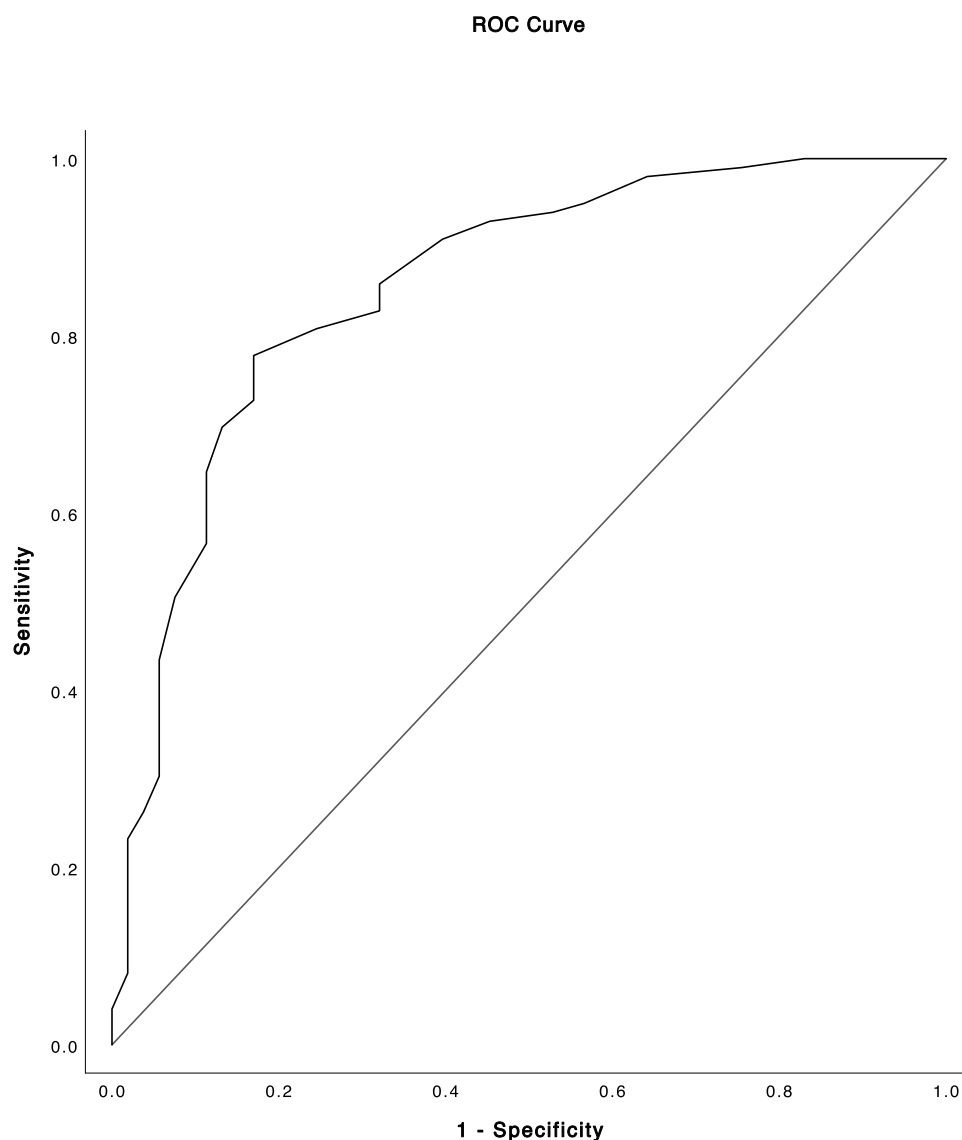
and 13% of adults with ADHD. As such, it could act as a valuable screening item that provides examiners with additional context within which to interpret the CII.

Our findings further point to the value of integrating multiple validity indicators that identify symptom overreport. Several such combinations may prove advantageous in a variety of assessment contexts, especially if additional validity indicators beyond those that use the original CAARS items are unavailable.

Validity indicators based on the detection of overreport appear to be particularly useful in referral contexts that require the highest possible sensitivity. With 50 to 73% of instructed simulators being identified, suspect *T*-score elevations on the DSM scales may serve as initial screening measures. In line with its intended purpose of detecting extreme subscale elevations (Suhr et al., 2011), the CII showed comparable sensitivity. However, suspect *T*-score elevations on individual DSM scales and suspect scores on the CII also occurred in our sample of adults with ADHD, resulting in low specificity. The ADHD Index (H) was among the least sensitive indicators under study and lacked specificity in our sample of adults with ADHD too. This finding stands in contrast to results described by Harrison and Armstrong (2016), who included the ADHD Index in their Exaggeration Index as instructed simulators commonly presented with elevated scores on this index.

Taking into consideration the number of scales or indices that show overreport may increase specificity where it is paramount. The co-occurrence of overreport on all three DSM scales showed a sensitivity of 45 to 46% and a specificity of 92% in the larger patient sample. Notably, 22% of patients with ADHD in the smaller subsample A-2 ($n = 4$) also overreported on all DSM scales, underscoring the importance of validating this marker in an independent sample. Suspect

Fig. 3 Receiver operating characteristics (ROC) curve indicating diagnostic accuracy of the new CII-ACI-Compound Index in identifying feigned ADHD (subset S-2 of *simulation group*, $n = 31$) relative to ADHD (subset A-2 of *ADHD group*, $n = 16$)
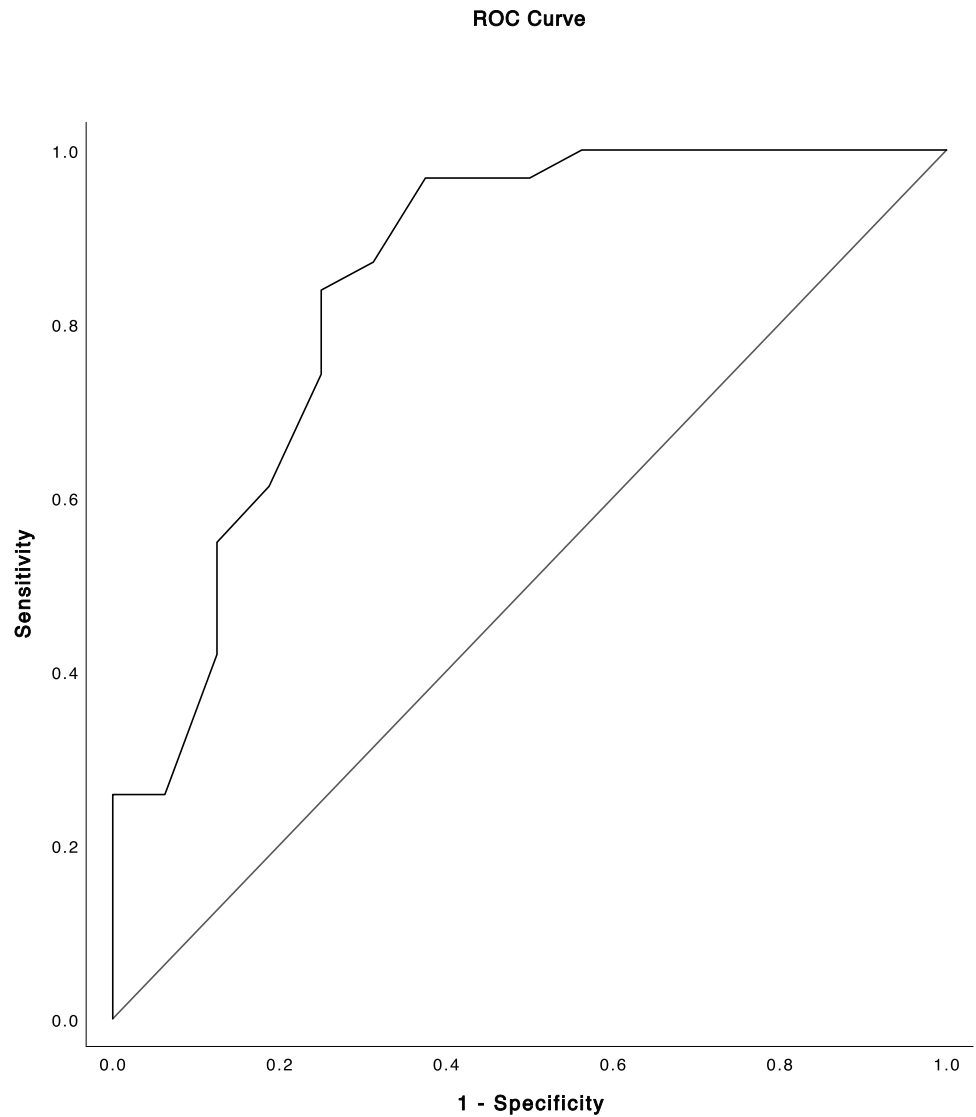

ROC Curve

Table 4 Percentage of participants identified by CAARS validity indicators

	Control group		ADHD group		Simulation group	
	Subset C-1	Subset C-2	Subset A-1	Subset A-2	Subset S-1	Subset S-2
T-score $\geq$ 80 on						
Any One DSM Scale	0.47	0.94	22.22	22.22	11.76	9.09
Any Two DSM Scales	0.47	0.47	29.63	11.11	20.59	12.12
All Three DSM Scales	0.16	0.00	7.41	22.22	46.08	45.45
DSM Inattention (E)	1.10	1.41	55.56	44.44	67.65	60.61
DSM Hyperactivity/ Impulsivity (F)	0.16	0.00	11.11	27.78	50.00	45.45
DSM Total (G)	0.63	0.47	37.04	38.89	73.53	63.64
ADHD Index (H)	0.16	0.47	16.67	21.43	15.31	9.68
CII	0.93	0.47	38.89	44.44	66.67	63.64
ACI	0.47	0.47	9.26	0.00	49.49	32.26

CII CAARS Infrequency Index (Suhr et al., 2011), *ACI* ADHD Credibility Index (Becke et al., 2021)

Table 5 Percentage of participants identified by configural features

	Control Group		ADHD Group		Simulation Group	
	Subset C-1	Subset C-2	Subset A-1	Subset A-2	Subset S-1	Subset S-2
Suspect results on						
CII and H	0.16	0.47	16.67	14.29	15.31	9.68
ACI and H	0.00	0.47	9.26	0.00	15.31	9.68
CII and ACI	0.47	0.47	9.26	0.00	43.88	29.03
All three DSM Scales and H	0.00	0.00	7.41	0.00	13.27	9.68
All three DSM Scales and CII	0.16	0.00	7.41	7.14	42.86	35.48
All three DSM Scales and ACI	0.16	0.00	5.56	0.00	30.61	16.13
All three DSM Scales, CII, and ACI	0.16	0.00	5.56	0.00	30.61	16.13
H or CII	0.94	0.47	38.89	56.25	67.33	63.64
H or ACI	0.63	0.47	16.67	28.57	50.00	32.26
CII or ACI	0.93	0.47	38.89	53.33	73.74	70.97
All three DSM Scales or H	0.31	0.47	16.67	43.75	48.51	45.45
All three DSM Scales or CII	0.94	0.47	38.89	44.44	69.61	69.69
All three DSM Scales or ACI	0.47	0.47	11.11	26.67	65.66	64.52
All three DSM Scales and (CII or ACI)	0.16	0.00	7.41	26.67	44.44	41.94

CII CAARS Infrequency Index (Suhr et al., 2011), *H* CAARS ADHD Index, *ACI* ADHD Credibility Index (Becke et al., 2021)

results on the CII rarely coincided with suspect *T*-score elevations on all DSM scales among adults with ADHD in either subsample. Yet this combination detected 35 to 43% of instructed simulators, making it another potentially useful marker of non-credible self-report.

Altogether, requiring multiple pre-existing validity indicators to yield suspect results for an examinee to be considered non-credible improved specificity significantly, but decreased sensitivity to feigned ADHD substantially. As expected, the specificity of such combinations was comparable to the highest specificity achieved by any single included indicator. Sensitivity was reduced to levels observed for the least sensitive indicator considered in combination. Jointly considering the CII and ACI, for example, raised the CII's specificity to adequate levels, but lowered sensitivity below true-positive rates noted for the ACI.

Classifying participants as non-credible if *either* one of multiple validity indicators showed a suspect result, on the other hand, led to high sensitivity and low specificity among adults with ADHD. Jointly interpreting the CII and ACI in this manner correctly identified 73% of instructed simulators

and therefore additional 7% of the *simulation group* that went undetected by the CII alone. However, nearly 40% of adults with ADHD were erroneously classified as non-credible. A different combination emerged as more useful: suspect scores on either the ACI *or* on all three DSM scales were more specific to feigned ADHD (89% specificity) while identifying ca. 65% of instructed simulators.

Cross-validation of these analyses using subsets 2 replicated many data patterns described for subsets 1, but overall classification accuracy was lower in the smaller subsamples. The CII-ACI-Compound Index showed a meaningful drop in both sensitivity and specificity when comparing the cross-validation with its initial results. Similarly, all but one configural feature under study (i.e., *T*-score elevations on all DSM scales *and* suspect result on the CII) demonstrated lower sensitivity in the cross-validation than they did initially. With a deviation of approximately 3%, the CII's sensitivity was most stable. Given the small sample size of 18 adults with ADHD, whose data were analyzed as part of the cross-validation, the variation in our results underscores the importance of their independent validation.

Table 6 Comparison of ROC results yielded by CII-ACI-Compound Index for subsets 1 and subsets 2

Comparison	Subsample	AUC	SE	*p*	95%-*CI*	
					Lower	Upper
ADHD Group vs. simulation group	Subsets 1	.855	.033	<.01	.790	.920
	Subsets 2	.847	.065	<.01	.719	.975
Unremarkable scores vs. overreport	Subsets 1	.785	.040	<.01	.706	.864
	Subsets 2	.612	.088	<.01	.440	.784

Limitations

The findings of this study ought to be seen in the light of limitations which may guide future research. First, differences in recruitment procedures led to significant group differences in demographics. The *simulation group* was significantly younger than the remaining groups, as it included many university students and thus a population highly relevant to research on feigned ADHD. Furthermore, the majority of instructed simulators identified as female, so the gender distribution within the *simulation group* differed meaningfully from gender-specific prevalence rates of ADHD. Should there be gender differences in the approaches to feigning (Heard, 2010) or the perception of ADHD, this imbalance introduced bias and reduced the representativeness of our sample. An additional concern regards the ratio of instructed simulators to adults with ADHD. The base rate of feigned ADHD was unrealistically high in our present sample, as the *simulation group* included more participants than the *ADHD group*. Positive and negative predictive values, which account for different base rates of feigned ADHD, indicate a drop in sensitivity of the CII-ACI-Compound Index in scenarios where base rates are lower than in our simulation design. Even so, the index remains useful at base rates approximating those seen in real-life (i.e., 20 to 50%; Marshall et al., 2010; Martin & Schroeder, 2020; Suhr et al., 2008; Sullivan et al., 2007). Generalization from instructed simulators to real-world feigning of adult ADHD is further limited by differences in incentives available to either group. The personal relevance or value of participants' chance at winning a tablet PC clearly differs from the incentives motivating people to feign ADHD in real life. Examiners not having been blinded to the instructions and incentives presented to the simulation group introduced another potential source of bias.

In contrast to the *simulation group*, which included mostly university-aged participants, the *control group* modeled a community sample and included comparable numbers of participants aged 25 years through 65 years. Collecting their data online allowed us to include a sample of participants from demographically and socioeconomically diverse backgrounds, but presented the challenge of lesser control over the circumstances under which they answered the questionnaires. Careless responding may be a common concern among control participants whose primary incentives for taking part in a study are monetary. We, therefore, excluded participants with suspiciously short response times (7.42% of respondents), yet other validity checks were not in place. As digitization progresses and electronic or online data collection becomes more common in research as well as healthcare, the quality and validity of data gathered in this manner ought to be thoroughly investigated. Their limitations should be weighed against those of convenience sampling or inclusion of participants recruited through researchers' contacts, which also present potential sources of biases.

Furthermore, the referral context of patients with ADHD and diagnostic procedure are associated with limitations that require thorough consideration. Data were collected in an outpatient clinic that offers comprehensive diagnostic check-ups to adults with suspected ADHD, who may or may not have received a formal diagnosis in childhood or adolescence. This recruitment procedure led to the selection of a patient sample that may not generalize to adults with ADHD who all have been formally diagnosed with ADHD in childhood. Symptom levels, clinically relevant impairments, and the clinical trajectory may differ between adults with childhood ADHD and adults with a first-time ADHD diagnosis. Even though our diagnostic procedure followed empirically informed guidelines of first-time adult ADHD (e.g., see Sibley, 2021), this diagnostic procedure bears risks of misdiagnosis. For example, false-positive diagnoses are likely to occur when ADHD symptoms are present in the absence of impairments, in over-demanding environments, or because of the confusion with stress responses, other mental conditions, effects of aging, or feigning. In contrast, false negatives are likely to occur in cases of symptom under-reporting, unestablished childhood difficulties, at referral for comorbidity, or in under-represented populations (Sibley, 2021). Thus, the sample characteristics of first-time adult ADHD diagnoses, and the associated threats to diagnostic veracity, may limit the implications of the data and may need replication on samples of adults with ADHD who have been formally diagnosed in childhood for the first time.

High rates of comorbidities observed in the present sample of adults with ADHD are typical for this referral context and population (Chen et al., 2018; Cumyn et al., 2009), especially if ADHD was unrecognized or inadequately treated prior to first-time diagnosis in adulthood. Even though experienced clinicians consulted various instruments and sources throughout the diagnostic process, diagnoses were not made without evidence of objective impairment, and all examinees included in the present study passed an independent performance validity test, inclusion of non-credible cases cannot be ruled out conclusively. Indeed, subsets of adults with ADHD included in the present study responded inconsistently or reported high levels of symptomatology in accordance with overreport. Lastly, patients' medication status at the time of their participation, which may have impacted recall of symptoms and their severity, could not be retrieved systematically.

The study design further lacked a clinical control group and a group of simulators instructed to feign general psychological distress or disorders other than ADHD. The question of whether the configural features presented here are specific to feigned ADHD, rather than general "faking bad" or overreport, therefore remains unanswered. Finally,

the CAARS-ACI administered to participants in the present study did not include items from the DES. We were consequently unable to consider the EI in our analyses. As the approach Harrison and colleagues (2016) followed in its development differed from all other available validity indicators, the EI would have been a useful addition to the instruments under investigation in this study.

Concluding Remarks

Incorporating five additional ACI items into the CAARS allows for the calculation of a CII-ACI-Compound Index, which offers a favorable trade-off between sensitivity and specificity, yet does not outperform the ACI significantly. Depending on the assessment context, different combinations of pre-existing validity indicators may present useful alternatives where the ACI is not available. Co-occurrence of suspect *T*-score elevations on all three DSM scales emerged as a potentially useful adjunct marker of feigned adult ADHD, which may warrant further investigation. The same held true for DSM scale elevations coinciding with suspect results on the CII. Combining multiple indicators in an "either/or" fashion may offer high sensitivity when screening for feigned ADHD. Examiners may, for example, consider elaborate validity testing for those who present with suspect results on the CII *or* overreport on all three DSM scales. Suspect scores on the ACI and overreport on all three DSM scales lend themselves to joint consideration in the same manner when screening for feigned ADHD.

Requiring multiple validity indicators to yield concurring, suspect results for an examinee to be considered noncredible increases specificity and may be appropriate in high-stakes settings. For instance, additional information provided by measures more sensitive to feigning than overreport (e.g., ACI or ACI item number 10) may improve the specificity of the CII when examining adults with pronounced symptomatology. Suspect results on the ACI coinciding with overreport on all three DSM scales are also uncommon among credible adults with ADHD and controls, and could warrant further validity testing.

Before finding widespread clinical application, these indicators or configural features should clearly undergo independent validation, ideally using known-group designs alongside simulation designs.

Declarations

References

Adler, L. A., Spencer, T., Faraone, S. V., Kessler, R. C., Howes, M. J., Biederman, J., & Secnik, K. (2006). Validity of pilot adult ADHD Self-Report Scale (ASRS) to rate adult ADHD symptoms. *Annals of Clinical Psychiatry, 18*(3), 145–148. https://doi.org/10.1080/10401230600801077

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (IV TR). *Washington, DC*. https://doi.org/10.1176/appi.books.9780890423349

American Psychiatric Association. (2013). DSM 5. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.books.9780890425596.744053

Barkley, R. A., & Murphy, K. R. (1998). *Attention-deficit hyperactivity disorder: A clinical workbook*. GUILFORD PUBLICATIONS INC.

Becke, M., Fuermaier, A. B. M., Buehren, J., Weisbrod, M., Aschenbrenner, S., Tucha, O., & Tucha, L. (2019). Utility of the Structured Interview of Reported Symptoms (SIRS-2) in detecting feigned adult attention-deficit/hyperactivity disorder. *Journal of Clinical and Experimental Neuropsychology, 41*(8), 786–802. https://doi.org/10.1080/13803395.2019.1621268

Becke, M., Tucha, L., Weisbrod, M., Aschenbrenner, S., Tucha, O., & Fuermaier, A. B. M. (2021). Non - credible symptom report in the clinical evaluation of adult ADHD : Development and initial validation of a new validity index embedded in the Conners ' adult ADHD rating scales. *Journal of Neural Transmission*. https://doi.org/10.1007/s00702-021-02318-y

Ben-Porath, Y., & Tellegen, A. (2008). *MMPI-2-RF: Minnesota Multiphasic Personality Inventory-2 Restructured Form*. Minneapolis: Pearson Assessment Systems.

Bernstein, E. M., & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous and Mental Disease*. https://doi.org/10.1097/00005053-198612000-00004

Biederman, J., Faraone, S. V., Spencer, T., Wilens, T., Norman, D., Lapey, K. A., & Doyle, A. (1993). Patterns of psychiatric comorbidity, cognition, and psychosocial functioning in adults with attention deficit hyperactivity disorder. *The American Journal of Psychiatry, 150*(12), 1792–1798. https://doi.org/10.1176/ajp.150.12.1792

Chen, Q., Hartman, C. A., Haavik, J., Harro, J., Klungsøyr, K., Hegvik, T.-A., & Larsson, H. (2018). Common psychiatric and metabolic comorbidity of adult attention-deficit/hyperactivity disorder: A population-based cross-sectional study. *PLoS ONE, 13*(9), e0204516–e0204516. https://doi.org/10.1371/journal.pone.0204516

Conners, C. K., Erhardt, D., & Sparrow, M. A. (1999). *Conners' Adult ADHD Rating Scales (CAARS)*. New York, NY, NY: Multihealth Systems.

Cook, C., Buelow, M. T., Lee, E., Howell, A., Morgan, B., Patel, K., … Suhr, J. (2017). Malingered attention deficit/hyperactivity disorder on the Conners' Adult ADHD Rating Scales. *Journal of Psychoeducational Assessment*, 073428291769693. https://doi.org/10.1177/0734282917696934

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.

Courrégé, S. C., Skeel, R. L., Feder, A. H., & Boress, K. S. (2019). The ADHD symptom infrequency scale (ASIS): A novel measure designed to detect adult ADHD simulators. *Psychological Assessment, 31*(7), 851–860. https://doi.org/10.1037/pas0000706

Cumyn, L., French, L., & Hechtman, L. (2009). Comorbidity in adults with attention-deficit hyperactivity disorder. *The Canadian Journal of Psychiatry, 54*(10), 673–683. https://doi.org/10.1177/070674370905401004

Edmundson, M., Berry, D. T. R., Combs, H. L., Brothers, S. L., Harp, J. P., Williams, A., & Scott, A. B. (2017). The effects of symptom information coaching on the feigning of adult ADHD. *Psychological Assessment, 29*(12), 1429–1436. https://doi.org/10.1037/pas0000478

Fuermaier, A. B. M., Tucha, O., Koerts, J., Lange, K. W., Weisbrod, M., Aschenbrenner, S., & Tucha, L. (2017a). Noncredible cognitive performance at clinical evaluation of adult ADHD: An embedded validity indicator in a visuospatial working memory test. *Psychological Assessment, 29*(12), 1466–1479. https://doi.org/10.1037/pas0000534

Fuermaier, A. B. M., Tucha, L., Koerts, J., Weisbrod, M., Grabemann, M., Zimmermann, M., & Tucha, O. (2016a). Evaluation of the CAARS Infrequency Index for the detection of noncredible ADHD symptom report in adulthood. *Journal of Psychoeducational Assessment, 34*(8), 739–750. https://doi.org/10.1177/0734282915626005

Fuermaier, A. B. M., Tucha, L., Koerts, J., Aschenbrenner, S., & Tucha, O. (2017b). *Groningen Effort Test (GET): Manual* (1st ed.). Schuhfried.

Fuermaier, A. B. M., Tucha, O., Koerts, J., Grabski, M., Lange, K. W., Weisbrod, M., & Tucha, L. (2016b). The development of an embedded figures test for the detection of feigned attention deficit hyperactivity disorder in adulthood. *PLoS ONE, 11*(10), e0164297. https://doi.org/10.1371/journal.pone.0164297

Fuermaier, Anselm B.M., Tucha, O., Koerts, J., Tucha, L., Thome, J., & Faltraco, F. (2021). Feigning ADHD and stimulant misuse among Dutch university students. *Journal of Neural Transmission*, (0123456789). https://doi.org/10.1007/s00702-020-02296-7

Greve, K. W., Bianchini, K. J., & Doane, B. M. (2006). Classification accuracy of the Test of Memory Malingering in traumatic brain injury: Results of a known-groups analysis. *Journal of clinical and experimental neuropsychology, 28*, 1176–1190.

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

Harrison, A. G., Edwards, M. J., & Parker, K. C. H. (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

Heard, K. V. (2010). Ethical and methodological considerations for gender researchers in forensic psychology. In J. C. Chisler & D. R. McCreary (Eds.), *Handbook of gender research in psychology - Volume 2: Gender research in social and applied psychology* (pp. 587–612). New York, NY: Springer Science + Business Media. https://doi.org/10.1007/978-1-4419-1467-5

Jachimowicz, G., & Geiselman, R. E. (2004). Comparison of ease of falsification of attention deficit hyperactivity disorder diagnosis using standard behavioral rating scales. *Cognitive Science*, *2*, 6–20. Retrieved from http://cogsci-online.ucsd.edu/2/2-1.pdf

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., & Walters, E. E. (2005). The World Health Organization adult ADHD self-report scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine, 35*(2), 245–256. https://doi.org/10.1017/S0033291704002892

Booksh, R. L., Pella, R. D., Singh, A. N., & Gouvier, W. D. (2010). Ability of college students to simulate ADHD on objective measures of attention. *Journal of Attention Disorders, 13*(4), 325–338. https://doi.org/10.1177/1087054708329927

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological assessment (4th ed.). Neuropsychological assessment (4th ed.).*

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

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*(6), 717–725. https://doi.org/10.1093/arclin/acaa017

Quinn, C., & a. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology, 18*(4), 379–395. https://doi.org/10.1016/S0887-6177(02)00150-6

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. (2008). An introduction to response styles. In *Clinical Assessment of Malingering and Deception* (3rd ed., pp. 3–13). New York, NY: GUILFORD PUBLICATIONS INC.

Rogers, R. (2018). Detection strategies for malingering and defensiveness. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 18–41). Guilford Press.

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

Sibley, M. H. (2021). Empirically-informed guidelines for first-time adult ADHD diagnosis. *Journal of Clinical and Experimental Neuropsychology, 43*(4), 340–351. https://doi.org/10.1080/13803395.2021.1923665

Smith, S. T., Cox, J., Mowle, E. N., & Edens, J. F. (2017). Intentional inattention: Detecting feigned attention-deficit/hyperactivity disorder on the personality assessment inventory. *Psychological Assessment, 29*(12), 1447–1457. https://doi.org/10.1037/pas0000435

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*(5), 521–530. https://doi.org/10.1016/j.acn.2008.05.003

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

Sullivan, G. M., & Feinn, R. (2012). Using effect size–Or why the P value is not enough. *Journal of Graduate Medical Education, 4*(3), 279–282. https://doi.org/10.4300/jgme-d-12-00156.1.

Tombaugh, T. N. (1996). *Test of Memory Malingering: TOMM*. Multihealth Systems.

Walls, B. D., Wallace, E. R., Brothers, S. L., & Berry, D. T. R. (2017). Utility of the Conners' Adult ADHD Rating Scale validity scales in identifying simulated attention-deficit hyperactivity disorder and random responding. *Psychological Assessment, 29*(12), 1437–1446. https://doi.org/10.1037/pas0000530

Ward, M. F., Wender, P. H., & Reimherr, F. W. (1993). The Wender Utah Rating-Sclae - An aid in the retrospective diagnosis of childhood attention-deficit hyperactivity disorder. *American Journal of Psychiatry, 150*(6), 885–890.