University of Groningen

Evaluation of the CAARS infrequency index for the detection of noncredible ADHD symptom report in adulthood.

Fuermaier, Anselm; Tucha, Lara; Koerts, Janneke; Weisbrod, Matthias; Grabemann, Marco; Zimmermann, Marco; Mette, Christian; Aschenbrenner, Steffen; Tucha, Oliver

Published in:
Journal of Psychoeducational Assessment

DOI:
[10.1177/0734282915626005](https://doi.org/10.1177/0734282915626005)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Fuermaier, A., Tucha, L., Koerts, J., Weisbrod, M., Grabemann, M., Zimmermann, M., Mette, C., Aschenbrenner, S., & Tucha, O. (2016). 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>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Evaluation of the CAARS Infrequency Index for the Detection of Noncredible ADHD Symptom Report in Adulthood

Journal of Psychoeducational Assessment
2016, Vol. 34(8) 739–750
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0734282915626005
jpa.sagepub.com

**Anselm B. M. Fuermaier¹, Lara Tucha¹, Janneke Koerts¹,
Matthias Weisbrod^{2,3}, Marco Grabemann⁴,
Marco Zimmermann⁴, Christian Mette⁴,
Steffen Aschenbrenner², and Oliver Tucha¹**

Abstract

The reliance on self-reports in detecting noncredible symptom report of attention-deficit/hyperactivity disorder in adulthood (aADHD) has been questioned due to findings showing that symptoms can easily be feigned on self-report scales. In response, Suhr and colleagues developed an infrequency index for the Conners' Adult ADHD Rating Scale (CII) and provided initial validation for its utility in detecting noncredible symptom report. The aim of this study was to evaluate the utility of the CII in detecting noncredible aADHD symptom report by using a simulation design. Data did not support the validity of the CII for the detection of noncredible aADHD symptoms, as it failed to differentiate instructed malingerers from genuine patients with sufficient accuracy. It is concluded that there is a need for infrequency scales composed of items that were specifically developed to be endorsed infrequently and embedded within valid self-report scales.

Keywords

adult ADHD, assessment, diagnosis, feigning, malingering, rating scale, self-report, infrequency scale

Patients' self-reports, as obtained by self-report symptom rating scales and clinical interviews, form one of the primary sources of information to establish the diagnosis of attention-deficit/hyperactivity disorder in adulthood (aADHD; American Psychiatric Association [APA], 2013; Kooij et al., 2008). However, the reliance on self-reports for establishing a valid diagnosis of aADHD has been questioned for several reasons, as (a) high base rates of ADHD symptoms are

¹University of Groningen, The Netherlands

²SRH Clinic Karlsbad-Langensteinbach, Germany

³University of Heidelberg, Germany

⁴University of Duisburg-Essen, Germany

Corresponding Author:

Anselm B. M. Fuermaier, Faculty of Behavioural and Social Sciences, Department of Clinical and Developmental Neuropsychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.
Email: a.b.m.fuermaier@rug.nl

endorsed by the general population and non-ADHD clinical populations (Murphy, Gordon, & Barkley, 2000; Suhr, Zimak, Buelow, & Fox, 2009), (b) “help-seeking behavior” of individuals may lead to overreporting of symptoms to convince and impress the examiner about the severity of their condition, and (c) individuals might be motivated by external incentives to deliberately feign aADHD, also described as malingering (APA, 2013). There are various external incentives discussed, which may motivate people to feign aADHD, including access to stimulant medication (Lensing, Zeiner, Sandvik, & Opjordsmoen, 2013), access to social welfare benefits, diminished criminal responsibility, or special academic accommodations (Tucha, Sontag, Walitza, & Lange, 2009). In fact, high base rates for noncredible performance and symptom reporting (ranging from 25% to 48%) have been estimated for college students presenting for clinical evaluation (Pella, Hill, Shelton, Elliott, & Gouvier, 2012; Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008; Sullivan, May, & Galbally, 2007).

Though these findings underline the urgent need for effective detection methods for noncredible effort and aADHD symptom reports, there is so far a surprising shortage of available research on the development and evaluation of such methods. While some methods showed promising classification accuracies in terms of sensitivity and specificity (e.g., Quinn, 2003), most studies concluded limited clinical utility of single measures in the identification of noncredible performance during ADHD evaluation, as only moderate sensitivity was reached (for comprehensive reviews on this topic, see Musso & Gouvier, 2014, as well as Tucha, Fuermaier, Koerts, Groen, & Thome, 2015). In response, Suhr, Buelow, and Riddle (2011) stressed the need for self-report instruments that include validity indices. They argued that such validity indices should consist of items that are embedded within valid and widely used self-report rating scales as this would make the detection strategy less obvious to individuals feigning the disorder (Suhr et al., 2011). Based on these considerations, Suhr and colleagues (2011) suggested a promising validity index that is embedded in the Conners’ Adult ADHD Rating Scale (CAARS; Conners, Erhardt, & Sparrow, 1998). This validity index, the Conners’ infrequency index (CII), was developed based on the assumption that individuals’ malingering symptoms are more likely to endorse symptoms that are infrequently endorsed by others, that is, typically developing individuals and genuine patients. In the development of this index, the authors analyzed a large number of CAARS responses and identified items that were endorsed infrequently by both typically developing individuals and genuine patients with aADHD. A CII score was obtained by summing up the responses of these infrequently endorsed items. The initial validation study showed good accuracy of the CII to predict noncredible high symptom report (78% to 92% overall accuracy) and moderate accuracy to predict noncredible performance on cognitive tests (67% overall accuracy; Suhr et al., 2011).

The development of the CII represents a promising contribution for the detection of noncredible aADHD symptom report. However, prior to its safe application in clinical practice, further validation of CII’s usefulness is necessary. The aim of the present study was, therefore, to explore the validity of the CII in detecting noncredible aADHD symptom report by applying a simulation design. In this study, typically developing participants were randomly allocated to either a control condition (instructions to perform tests to the best of their ability) or to one of four experimental simulation conditions. In the simulation conditions, participants were instructed to feign aADHD on the CAARS. In three of the conditions, participants were coached prior to simulation by informing about diagnostic criteria of ADHD and/or subscales of ADHD rating scales. This approach was chosen to provide instructed simulators with similar information as individuals attempting to feign aADHD in clinical setting may acquire when preparing themselves. Furthermore, CAARS responses were collected from a group of genuine patients with aADHD. A comparison of responses between instructed malingerers and genuine patients provides a controlled experimental evaluation of the utility of the CII in detecting noncredible aADHD symptom report. The utility would be supported if results showed that the CII can also differentiate well-prepared (coached) instructed simulators from individuals suffering from aADHD.

Table 1. Characteristics of Participants ($M \pm SD$).

	CG	NSG	SSG	TSG	STSG	aADHD
N	82	87	91	86	89	52
Age (in years)	27.2 $\pm$ 10.7	28.2 $\pm$ 11.6	27.3 $\pm$ 10.4	28.1 $\pm$ 11.3	27.2 $\pm$ 10.9	29.6 $\pm$ 7.7
Gender (female/male)	46/36	45/42	57/34	54/32	44/45	24/28
Intellectual functions (IQ)	102.0 $\pm$ 12.1	101.8 $\pm$ 10.8	101.3 $\pm$ 9.8	100.4 $\pm$ 8.7	102.9 $\pm$ 13.9	102.6 $\pm$ 8.8
WURS-K	12.5 $\pm$ 9.2	14.4 $\pm$ 10.9	12.3 $\pm$ 10.4	12.5 $\pm$ 10.5	12.7 $\pm$ 9.9	38.9 $\pm$ 10.5 ^{a,b,c,d,e}
ADHD self-report scale	10.2 $\pm$ 5.7	10.3 $\pm$ 6.1	10.1 $\pm$ 6.4	9.5 $\pm$ 5.1	10.4 $\pm$ 7.0	31.1 $\pm$ 8.6 ^{a,b,c,d,e}
ADHD knowledge	15.3 $\pm$ 6.9	16.3 $\pm$ 6.9	18.6 $\pm$ 6 ^a	17.4 $\pm$ 5.9	19.1 $\pm$ 6.5 ^a	—

Note. Age, sex, IQ, and ADHD symptom severity data were obtained at the beginning of the experiment from all participants. CG completed the ADHD knowledge questionnaire directly after descriptive information has been obtained. Simulation groups completed the ADHD knowledge questionnaire after simulation instructions have been studied, but before participants started to feign aADHD. aADHD = attention-deficit/hyperactivity disorder in adulthood; CG = control group; NSG = naive simulation group; SSG = symptom-coached simulation group; TSG = test-coached simulation group; STSG = symptom- and test-coached simulation group; WURS-K = Wender Utah Rating Scale—short version.

^a $p < .05$ when compared with CG.

^b $p < .05$ when compared with NSG.

^c $p < .05$ when compared with SSG.

^d $p < .05$ when compared with TSG.

^e $p < .05$ when compared with STSG.

Method

Participants

Patients with aADHD. Fifty-two patients with aADHD participated in the study (Table 1). Patients were self-referred or referred from local psychiatrists or neurologists in the Departments of Psychiatry and Psychotherapy of the SRH Clinic Karlsbad-Langensteinbach or LVR-Hospital Essen, Germany. All patients with aADHD were invited to take part in the study on a voluntary basis. A diagnostic assessment for aADHD as well as a participation in the research project was offered to all participants. In the diagnostic assessment of the 52 patients with aADHD, 42 patients met *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; APA, 1994) criteria for aADHD—combined type and 10 patients met criteria for aADHD—predominantly inattentive type. None of the patients met criteria for aADHD—hyperactive-impulsive type. The overrepresentation of patients with ADHD of the combined subtype supports the argument that patients of the combined type are more likely to be referred for clinical evaluation (Willcutt, 2012). Eight of the 52 patients with aADHD were diagnosed with one or more comorbid disorders, including mood disorders ($n = 6$), anxiety disorder ($n = 1$), and personality disorders ($n = 2$). Furthermore, patients have not been treated for their ADHD symptoms, either by pharmacological or by nonpharmacological approaches. However, two patients with aADHD were treated with antidepressant medication because of comorbid disorders at the time of the study.

Typically developing individuals. Four hundred thirty-five typically developing individuals participated in the study. Characteristics of participants, including age, gender, intellectual functions, and self-reported ADHD symptom severity, are presented in Table 1. Typically developing individuals were recruited via public announcements, word of mouth, and through contacts of the researchers involved. All participants were invited to take part in the study on a voluntary basis. Two hundred forty-six typically developing individuals were females and 189 were males. Typically developing participants' age ranged from 18 to 59 years, with a mean age of 27.6 ± 11.0 years, and a mean IQ of 101.7 ± 11.2 points. None of the typically developing individuals reported to have a history of neurological or psychiatric disease and none were taking any medication known to affect the central nervous system. ADHD symptom severity of typically developing

individuals was measured with two standardized self-report rating scales designed to quantify current and retrospective ADHD symptoms (Rösler, Retz-Junginger, Retz, & Stieglitz, 2008). Scores of all typically developing participants were below the cutoff value suggesting a clinical level of ADHD symptom severity. Typically developing participants were randomly assigned to one of five groups, that is, the *control group* (CG), the *naive simulation group* (NSG), the *symptom-coached simulation group* (SSG), the *test-coached simulation group* (TSG), and the *symptom- and test-coached simulation group* (STSG). The groups differed with regard to instruction and information they were provided prior to the assessment. This is relevant to increase external validity of this study, as individuals attempting to feign aADHD in clinical setting likely prepare themselves before diagnostic evaluation by informing about symptom criteria of ADHD as outlined in the *DSM*, and/or information about instruments used at a typical diagnostic assessment. The information provided to the different groups of the present study is thoroughly described below in the "Procedure" section. With regard to demographic characteristics (Table 1), groups did not differ significantly with regard to age, $F(5, 481) = 0.453$; $p = .811$, gender, $\chi^2(5) = 7.403$; $p = .192$, and intellectual functions, $F(5, 471) = 0.529$; $p = .754$. This is important to note as differences in demographic variables might affect the ability to simulate aADHD. However, as expected, significant group differences were obtained with regard to self-reported symptom severity of both current ADHD symptoms, $F(5, 465) = 72.60$; $p < .001$, and retrospective ADHD symptoms, $F(5, 465) = 42.35$; $p < .001$. Patients with aADHD scored significantly higher on both self-report rating scales than all groups of typically developing individuals ($p < .001$ for all comparisons). Nonsignificant differences between simulation groups in these characteristics are important to control for confounding variables that may affect simulation abilities. To check whether the different information provided had an effect on participants' knowledge about ADHD, participants' knowledge was assessed via a brief ADHD knowledge questionnaire (see material and procedure sections for more detailed information). Group comparison indicated that these five groups indeed differed in their knowledge about ADHD, $F(4, 414) = 4.870$, $p = .001$. Participants in the CG had significantly less knowledge about ADHD than participants in the SSG ($p = .030$) and STSG ($p = .008$).

Materials

Intellectual functions. Intellectual functions (IQ) were measured using the Multiple Choice Vocabulary Test (Lehrl, 1995; Lehrl, Triebig, & Fischer, 1995; Suslow, 2009; see supplementary file). This test was included to ensure that simulation groups did not differ significantly with regard to intellectual functioning. It must be noted that the instruction to feign aADHD requires the participants to assume the role of someone with aADHD. The performance of successfully feigning aADHD in a simulation design might therefore be affected by individual characteristics such as intellectual functions.

Clinical assessment of ADHD symptom severity. Retrospective and current ADHD symptom severity was assessed with the short version of the Wender Utah Rating Scale (WURS-K) and the ADHD self-report scale (Rösler et al., 2008; Ward, Wender, & Reimherr, 1993; see supplementary file). These scales were used to support the diagnostic process of patients with ADHD by indicating clinical levels of ADHD symptom severity.

ADHD knowledge questionnaire. Participants' knowledge about ADHD was assessed by a self-developed questionnaire consisting of 39 statements about ADHD (see supplementary file).

CAARS. The CAARS (self-report, long version; Conners et al., 1998) is a 66-item inventory that addresses self-reported aADHD symptoms (Christiansen et al., 2012; Erhardt, Epstein, Conners,

Parker, & Sitarenios, 1999). Answers are scored on a 4-point scale (0 = *not at all, never*; 1 = *just a little, once in a while*; 2 = *pretty much, often*; 3 = *very much, very frequently*). Participants were eventually informed of omitted items and were given the chance to complete to avoid missing values. Scores on individual items are summed up yielding eight different scales, with some items contributing to more than one scale. In addition, a validity index is obtained (CAARS-inconsistency) by summing up the numeric difference of scores on eight pairs of items that measure similar content. Furthermore, and in addition to these classic scales, the CII as suggested by Suhr and colleagues (Suhr et al., 2011) is calculated by summing up the scores of Items 21, 22, 23, 26, 30, 34, 43, 45, 49, 51, 52, and 62. For the purpose of this study and thereby following the work of Suhr and colleagues (2011), two classic scales of the CAARS (CAARS-E [*DSM-IV* inattention symptoms] and CAARS-F [*DSM-IV* hyperactivity/impulsivity symptoms]) as well as the CII were used to explore their utility to predict noncredible aADHD symptom report. Furthermore, the CAARS-inconsistency was included aiming to identify instructed simulators with inconsistent strategies used to feigning ADHD. CAARS scores were used for research purposes but not to describe clinical levels of ADHD symptom severity of patients with ADHD.

Design and Procedure

Assessment of patients with aADHD. Patients with aADHD were assessed individually and received no reward for participation. Written informed consent was sought from all participants prior to the assessment. It was pointed out to patients that all data collected in the research project will be analyzed anonymously and will not affect clinical assessment and treatment. Diagnostic assessments for aADHD were performed by experienced clinicians associated with the Departments of Psychiatry and Psychotherapy and involved a clinical psychiatric interview according to *DSM-IV* criteria for ADHD as devised by Barkley and Murphy (1998) including the retrospective diagnosis of ADHD in childhood (primary school age) and current symptoms. All diagnoses were made on mutual agreement between at least two clinicians belonging to a diagnostic team experienced in the assessment and treatment of adults with ADHD. Clinicians of this diagnostic team were licensed psychologists and had an explicit research interest in adult ADHD as evidenced by contributions to scientific publications. Patients were only included if clinicians unanimously agreed upon the presence of an aADHD diagnosis. The diagnostic assessment also included the identification of objective impairments supporting the diagnosis of ADHD (e.g., school reports, failure in academic and/or occupational achievement) and comprised, if possible, multiple informants, such as employer evaluation, partner reports, or parent reports. Moreover, all participants completed two standardized self-report rating scales designed to quantify current and retrospective ADHD symptoms (WURS-K and ADHD self-report scale). Self-reported ADHD symptom severity on these two scales was used for clinical purposes to support the diagnostic process. All patients scored above the recommended cutoffs on these scales, indicating clinical relevant ADHD symptom severity retrospectively for childhood and currently. Moreover, patients with aADHD were asked to fill out the CAARS to the best of their knowledge and not to seek help from the examiner or to discuss questions or their responses. The CAARS was used for research purposes only but not to describe clinical levels of ADHD symptom severity in the diagnostic process. The completion of the CAARS was placed at the beginning of a larger assessment, which took about 2.5 hr in total. The study was conducted in compliance with ethical standards and was approved by local institutional ethical committees.

Assessment of typically developing participants. All typically developing participants gave written informed consent for the participation in the study and were assessed individually by an examiner in a quiet environment. At the beginning of the experiment, descriptive and anamnestic information was obtained from all participants, including age, sex, intellectual functions, and

self-reported ADHD symptom severity. Furthermore, participants were asked for any history of psychiatric or neurological diseases as well as pharmacological treatment. After descriptive information has been obtained, participants were randomly assigned to one of five experimental groups, that is, the CG, NSG, SSG, TSG, or the STSG. The subsequent assessment procedure differed between participants of the various experimental groups.

CG. After completion of descriptive information and group assignment, participants of the CG were first asked to complete the ADHD knowledge questionnaire and then asked to fill in the CAARS to the best of their knowledge and not to seek help from the examiner or to discuss questions or their responses. The duration of the assessment for the CG was about 40 min.

Simulation groups. After completion of descriptive information and group assignment, participants of the simulation groups (NSG, SSG, TSG, and STSG) were presented a scenario that introduced several benefits that may come with a diagnosis of aADHD (see supplementary file). Information provided in the scenario was restricted to support participants to assume their role but did not contain information about the symptoms or nature of aADHD. To encourage participants to feign aADHD in a believable and realistic manner, participants were informed that the participant who feigns the condition best would be awarded with a recent tablet personal computer (PC). However, the tablet PC was in fact assigned randomly to one of the participants across all conditions due to ethical reasons. After the task of feigning aADHD had been introduced to all participants of the simulation groups, procedures differed with regard to the additional information participants received. The NSG received no further information and no suggestions of how to fake aADHD on the CAARS. The SSG received a description of the diagnostic criteria for ADHD as outlined in the *DSM-IV* (see supplementary file). This approach has been shown in previous studies to provide instructed malingerers with sufficient information to become familiar with the characteristics of ADHD (Tucha et al., 2009). In the TSG, participants were informed about how a diagnostic assessment of aADHD is commonly performed, including the use of a clinical interview and ADHD symptom rating scales (i.e., the CAARS; see supplementary file). The STSG received both the information about the diagnostic criteria of ADHD as outlined in the *DSM-IV*, as well as the information about the diagnostic assessment for aADHD. After having received and read the respective information for simulation, participants were requested to respond to a number of questions on the content of information to ascertain that they read and understood the information provided. Participants were then asked to complete the ADHD knowledge questionnaire, after having studied simulation instructions, but before starting to feign aADHD. Finally, participants were requested to start feigning aADHD and to complete the CAARS so as to appear as if they would suffer from aADHD. At the end of the assessment, participants were instructed to stop feigning aADHD and were debriefed. The duration of the assessment was about 60 min in total.

Statistical Analysis

MANOVA was calculated to compare groups on all scales of the CAARS in a common analysis. Univariate ANOVA were performed to compare groups on individual scales in separate analyses, whereas post hoc tests (Scheffé) were used for specific pairwise comparisons of interest, that is, between patients and control participants, as well as between patients and simulation groups. Effect sizes (Cohen's *d*) were calculated to indicate the magnitude of group differences.

Moreover, binary logistic regression models were calculated to determine the validity of CAARS scales in predicting malingered aADHD relative to genuine patients. Significance tests of regression models indicate the usefulness of the scales compared with chance level, whereas a comparison of R^2 values reveals the predictive value of the CII to detect feigned aADHD relative

to classic scales. Logistic regression models were calculated for the simulation groups NSG and STSG, to explore the utility to detect feigned aADHD relative to a naive group (effectiveness of simulation design in which no further information is presented), as well as relative to a well-prepared group as it is likely to be seen by prepared individuals attempting to feign aADHD in clinical practice. The predictive validity of CAARS scales were tested both in separate analyses by multiple binary logistic regression models, as well as by hierarchical binary logistic regression models in which classic scales of the CAARS were entered to the model in Block 1 and the CII in Block 2. This approach was chosen to evaluate whether the introduction of the CII adds predictive value to detect noncredible aADHD symptom report assuming the presence of classic scales of the CAARS. Furthermore, and most importantly for clinical interpretation, receiver operating characteristic (ROC) analyses were conducted to explore the accuracy of the CAARS scales in detecting individuals malingering aADHD (NSG or STSG, respectively) relative to genuine patients. ROC analysis allows for determination of an overall accuracy of classification, as well as classification statistics to address specific goals, that is, high sensitivity or high specificity. An inspection of classification statistics provides direct information with regard to the utility to detect noncredible symptom report. For example, it was suggested that tests used to detect noncredible symptom report should achieve specificity to genuine aADHD of at least 90% (Boone, 2007).

Results

Group Comparisons

CAARS scores of patients with aADHD, control participants, and simulation groups are presented in Table 2. MANOVA revealed a significant difference in CAARS scores between groups of large size, Wilks's lambda = 0.430, $F(20, 1586) = 22.99$, $p < .001$, $\eta^2 = 0.190$. Univariate comparisons indicated significant group differences on all four scales, that is, CAARS-E, $F(5, 481) = 87.54$, $p < .001$, $\eta^2 = 0.476$, CAARS-F, $F(5, 481) = 81.97$, $p < .001$, $\eta^2 = 0.460$, CAARS-inconsistency, $F(5, 481) = 4.29$, $p = .001$, $\eta^2 = 0.043$, and CII, $F(5, 481) = 56.35$, $p < .001$, $\eta^2 = 0.369$. Post hoc pairwise comparisons (Scheffé) were performed to compare patients with the CG as well as patients with simulation groups. Post hoc tests showed that patients with aADHD, as expected, reported higher ADHD symptom severity than the CG on CAARS-E ($p < .001$, $d = 1.86$) and CAARS-F ($p < .001$, $d = 1.34$). Significant higher scores of patients with aADHD compared with the CG was also found on CAARS-inconsistency ($p = .007$, $d = 0.62$) and CII ($p < .001$, $d = 1.53$). Comparing patients with simulation groups revealed overestimated symptom severity reported by instructed malingerers. This was indicated by group comparisons showing that the NSG scored significantly higher than genuine patients on CAARS-F ($p < .001$, $d = 1.25$) and CII ($p = .025$, $d = 0.62$), whereas no significant differences were found on CAARS-E ($p = .248$, $d = 0.42$) and CAARS-inconsistency ($p = .914$, $d = 0.18$). The SSG revealed significantly higher scores than genuine patients on CAARS-F ($p < .001$, $d = 0.87$) and CAARS-E ($p = .020$, $d = 0.63$), with no significant differences observed on CAARS-inconsistency ($p = .869$, $d = 0.22$) and CII ($p = .228$, $d = 0.44$). The TSG was found to score significantly higher than genuine patients on CAARS-F ($p < .001$, $d = 1.26$), CAARS-E ($p = .009$, $d = 0.71$), and CII ($p = .044$, $d = 0.59$), whereas the difference in CAARS-inconsistency did not reach significance ($p = .414$, $d = 0.33$). The STSG, finally, scored significantly higher than genuine patients on CAARS-F ($p = .016$, $d = 0.56$), whereas nonsignificant differences were observed on CAARS-E ($p = .668$, $d = 0.27$), CAARS-inconsistency ($p = .211$, $d = 0.46$), and CII ($p = .995$, $d = 0.10$).

Prediction of Group Membership

On the basis of a collapsed group of NSG and genuine patients, binary logistic regression models showed that CAARS-E, CAARS-F, and CII significantly predict malingerers (Table 3). However,

Table 2. CAARS Scores ($M \pm SD$) of Participants.

	CG	NSG	SSG	TSG	STSG	aADHD
CAARS-E	6.7 $\pm$ 5.3	19.0 $\pm$ 4.8	20.0 $\pm$ 4.6	20.2 $\pm$ 4.1	18.3 $\pm$ 5.4	16.8 $\pm$ 5.8 ^{a,b,c}
CAARS-F	5.6 $\pm$ 5.4	19.5 $\pm$ 4.6	18.0 $\pm$ 5.5	19.3 $\pm$ 4.2	16.6 $\pm$ 6.4	13.1 $\pm$ 5.9 ^{a,b,c,d,e}
CAARS-inconsistency	4.8 $\pm$ 2.4	5.9 $\pm$ 2.1	5.8 $\pm$ 2.1	5.5 $\pm$ 2.4	5.3 $\pm$ 2.0	6.3 $\pm$ 2.5 ^a
CII	7.0 $\pm$ 5.7	20.6 $\pm$ 6.1	19.5 $\pm$ 6.2	20.3 $\pm$ 5.8	17.3 $\pm$ 6.5	16.6 $\pm$ 7.1 ^{a,c,d}

Note. aADHD = attention-deficit/hyperactivity disorder in adulthood; CAARS = Conners' Adult ADHD Rating Scale; CG = control group; NSG = naive simulation group; SSG = symptom-coached simulation group; TSG = test-coached simulation group; STSG = symptom- and test-coached simulation group; CAARS-E = CAARS *DSM-IV* inattention symptoms; CAARS-F = CAARS *DSM-IV* hyperactivity/impulsivity symptoms; CII = CAARS infrequency index.

Pairwise comparisons between aADHD and all groups of typically developing individuals:

^a $p < .05$ when compared with CG.

^b $p < .05$ when compared with SSG.

^c $p < .05$ when compared with TSG.

^d $p < .05$ when compared with NSG.

^e $p < .05$ when compared with STSG.

the predictive validity of CAARS-F (Cox and Snell $R^2 = 26.1\%$) was considerably higher than the predictive validity of CII (Cox and Snell $R^2 = 8.0\%$) and CAARS-E (Cox and Snell $R^2 = 4.2\%$). CAARS-inconsistency (Cox and Snell $R^2 = 1.0\%$) did not significantly predict individuals of the NSG relative to genuine patients. A hierarchical binary logistic regression analysis demonstrated that the CII did not add significant predictive validity in detecting malingering (NSG) if classic scales of the CAARS (CAARS-E, CAARS-F, CAARS-inconsistency) were included first to the model. Most importantly for clinical interpretation, ROC analyses showed that the accuracy of CAARS scores in detecting malingering (NSG) relative to genuine patients was higher for CAARS-F (area under the curve [AUC] = 81.9%; $SE = 0.039$; $p < .001$) than for CII (AUC = 65.8%; $SE = 0.050$; $p = .002$). Lower classification rates were found for CAARS-E (AUC = 59.8%; $SE = 0.049$; $p = .054$) and CAARS-inconsistency (AUC = 55.1%; $SE = 0.052$; $p = .315$). Application of the cutoff of 21 or higher, as suggested by Suhr and colleagues (2011) demonstrated unsatisfactory classification rates, that is, sensitivity (52%), specificity (65%), positive predictive power (71%), and negative predictive power (45 %).

With regard to a collapsed group of STSG and genuine patients, binary logistic regression models showed that CAARS-E (Cox and Snell $R^2 = 1.8\%$), CAARS-F (Cox and Snell $R^2 = 7.0\%$), and CAARS-inconsistency (Cox and Snell $R^2 = 4.9\%$) were all superior to CII (Cox and Snell $R^2 = 0.2\%$) in predicting malingered aADHD (STSG; Table 3). A hierarchical binary logistic regression model demonstrated that the predictive validity of classic scales of the CAARS (CAARS-E, CAARS-F, CAARS-inconsistency) in detecting malingering (STSG) was increased by 5.3% additional explained variance in Block 2 when CII was added to the model. Most convincing results for clinical use were given by ROC analyses, demonstrating that the accuracy of CAARS scales in detecting malingered aADHD (STSG) was low to moderate for all scales, whereas CAARS-F (AUC = 67.8%; $SE = .047$; $p < .001$) was superior to CAARS-E (AUC = 58.1%; $SE = 0.050$; $p = .111$), CAARS-inconsistency (AUC = 55.1%; $SE = .052$; $p = .315$), and CII (AUC = 53.8%; $SE = .052$; $p = .453$). The suggested cutoff of 21 or higher on the CII for noncredible symptom report (Suhr et al., 2011) demonstrated unsatisfactory classification rates, that is, sensitivity (32%), specificity (65%), positive predictive power (61%), and negative predictive power (36%).

Discussion

Simulation conditions of the present study comprised the instruction to feign aADHD and to complete the CAARS so as to appear to suffer from aADHD. There is good indication that group

Table 3. Binary Logistic Regression Models of the CAARS to Predict Feigned aADHD.

Univariate binary logistic regression models				
Scale	Comparison	$\chi^2(1)$	<i>p</i> value	<i>R</i> ^{2a}
CAARS-E	aADHD—NSG	5.90	.015*	.042
CAARS-E	aADHD—STSG	2.53	.112	.018
CAARS-F	aADHD—NSG	42.06	<.001*	.261
CAARS-F	aADHD—STSG	10.29	.001*	.070
CAARS-inconsistency	aADHD—NSG	1.45	.228	.010
CAARS-inconsistency	aADHD—STSG	7.11	.008*	.049
CII	aADHD—NSG	11.53	.001*	.080
CII	aADHD—STSG	0.35	.555	.002
Hierarchical binary logistic regression models				
	Comparison	$\chi^2(df)$	<i>p</i> value	<i>R</i> ^{2a}
Model 1				
Block 1 ^b	aADHD—NSG	45.16 (3)	<.001*	.277
Block 2 ^c	aADHD—NSG	1.38 (4)	.241	.284
Model 2				
Block 1 ^b	aADHD—STSG	16.99 (3)	.001*	.114
Block 2 ^c	aADHD—STSG	25.79 (4)	<.001*	.167

Note. aADHD = attention-deficit/hyperactivity disorder in adulthood; CAARS = Conners' Adult ADHD Rating Scale; CAARS-E = CAARS DSM-IV inattention symptoms; CAARS-F = CAARS DSM-IV hyperactivity/impulsivity symptoms; CII = CAARS infrequency index; NSG = naive simulation group; STSG = symptom- and test-coached simulation group.

^aCox and Snell *R*².

^bIncluded variables in Block 1: CAARS-E, CAARS-F, CAARS-inconsistency.

^cIncluded variables in Block 2: CII.

*Significant at *p* < .05.

manipulations were successful as all participants were able to respond to questions regarding simulation instructions and also indicated after completion of the study that they have followed instructions. Further evidence is given by group statistics demonstrating clearly elevated ADHD symptom scores by simulation groups compared with the CG.

Group comparisons between instructed malingerers and genuine patients revealed that individuals instructed to feign aADHD largely overestimated symptom severity as experienced by patients with aADHD, as shown by significant differences reaching effect sizes of up to more than one standard deviation ($d > 1.0$) on CAARS-E, CAARS-F, and CII. However, data analysis does not support the utility of the CII for the detection of noncredible aADHD symptom report. The limited clinical utility is most clearly revealed in ROC analyses, demonstrating insufficient overall accuracy (53.8% or 65.8%), as well as unsatisfactory classification rates in terms of sensitivity (32% or 52%) and specificity (65%) when applying a cutoff of 21 or higher for noncredible symptom report (Suhr et al., 2011). Even though significant differences of CII scores were found between simulation groups (NSG and TSG) and genuine patients, group differences were not large enough to be of clinical utility, resulting in insufficient classification rates of even unprepared instructed malingerers (sensitivity 52%, specificity 65%) relative to genuine patients. These insufficient group differences may partly be explained by relatively high CII scores of the present patient group, that differed from the CG by more than 1.5 standard deviations ($d > 1.5$) and may thus not fulfill the criterion of an infrequency scale containing items that are infrequently endorsed

by most individuals. Comparing the CII with classic scales of the CAARS, it was shown that the CII did neither outperform classic scales in the ability to detect malingered aADHD, nor did the CII add considerable predictive power to classic scales in hierarchical analyses. However, with regard to hierarchical analyses, it must be noted that three items used to calculate the CII are derived from CAARS-F, which may affect incremental prediction. On the basis of the present data, it must therefore be concluded that the promising results of Suhr and colleagues (2011) regarding the use of the CII cannot be confirmed.

These results stimulate discussions on the utility of the long-lasting tradition of infrequency scales included in measures of psychopathology. It can be concluded that infrequency scales do not represent an appropriate mean for the detection of noncredible aADHD symptom report, as the present analysis demonstrated that even unprepared instructed malingerers (NSG) could not be sufficiently differentiated from genuine patients. In contrast to this view, it has been speculated (Suhr et al., 2011) that infrequency scales may in fact be an appropriate mean for the detection of noncredible aADHD symptom report. However, useful and valid infrequency scales for the detection of noncredible aADHD symptom report would be composed by items that were specifically developed to be endorsed infrequently, rather than by items that indeed describe common symptoms of ADHD but that have only infrequently been selected by a specific patient sample. Thus, the merely statistically identified items of the CAARS that constitute the CII do not represent a classic infrequency scale.

It must be noted that recent research consistently emphasized that the most helpful and sensitive approach for the detection of noncredible effort and aADHD symptom reporting was a combination of tests that were originally developed for the detection of noncredible performance (Musso & Gouvier, 2014; Tucha et al., 2015). However, such measures specifically designed for the detection of feigned aADHD are still lacking as most of the tests available were designed for detecting noncredible cognitive dysfunctions following acquired brain damage. Consequently, a promising route to help clinicians detecting noncredible aADHD symptom reports might be the development of infrequency scales consisting of items that were originally developed to be endorsed infrequently by patients with aADHD, as well as performance tests specifically developed for the detection of noncredible cognitive dysfunctions associated with ADHD.

As a limitation of the present study, the clinical assessment of patients with aADHD failed to include a validity measure indicating noncredible symptom reporting. Thus, it cannot be excluded that a small proportion of the sample of patients with aADHD may have shown noncredible symptom report (questioning diagnostic veracity), which may have resulted in an overreporting response bias and may have thus confounded the present analyses. To prevent the inclusion of noncredible individuals to the present patient sample, all diagnoses were made on agreement between at least two experienced clinicians. A conservative approach was employed by including patients only if clinicians unanimously agreed upon the presence of an aADHD diagnosis. Furthermore, the diagnostic assessment included the identification of objective impairments supporting the diagnosis of ADHD and comprised multiple informants although this information was not assessed on a common metric. Furthermore, as another limitation, it must be noted that the significant higher CAARS-inconsistency scores of patients with ADHD compared with the CG may indicate invalid symptom reporting. As an alternative explanation, this could also be explained by inattentive symptoms of patients (e.g., frequently reported lapses of attention), which may have resulted in problems responding consistently throughout a large number of items.

Coaching of participants prior to simulation is an important characteristic of studies using simulation designs as it creates more realistic scenarios by informing and preparing participants to feign the condition more successfully during the assessment. There are indications assuming that coaching of participants with regard to symptoms and tests (STSG) was successful, as the STSG (a) scored significantly higher on the ADHD knowledge questionnaire than the CG and (b)

scored closer to genuine patients on the CAARS symptom severity as compared with the NSG. However, only marginal differences between NSG, SSG, and TSG stimulate discussions on the effects of coaching in simulation designs. Previous research on the detection of noncredible performance and symptom report following brain injury has already pointed out that coaching of instructed malingerers might not affect simulation success considerably (Dunn, Shear, Howe, & Ris, 2003). Simulation designs in general can be criticized for a limited external validity, as the motive and incentive to malingering aADHD in an experimental setting, of course, does not match the setting in real life situations (Rogers, Harrell, & Liff, 1993). The validity of conclusions drawn from studies employing simulation designs would therefore benefit from confirmation of studies using different research designs, for example, *known-groups comparisons*.

Of note, the initial validation study of the CII (Suhr et al., 2011) was based on a sample of college students diagnosed with aADHD, thus indicating potential utility of the CII for detecting noncredible symptom report in the academic context. In the present study, however, referred patients from local psychiatrists or neurologists were assessed, so that the current conclusions are based on a different group of aADHD patients. Even though the present study did not provide support for the validity of the CII, the CII may still have predictive utility for the detection of noncredible aADHD symptom report among college students. Further validation studies are needed to address this issue, for example, by applying simulation designs on samples of college students. These studies should also include other effort tests to compare their utility to detect noncredible aADHD symptom report with the CII.

Declaration of Conflicting Interests

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

Funding

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

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric.
- Barkley, R. A., & Murphy, K. R. (1998). *Attention-deficit hyperactivity disorder: A clinical workbook*. New York, NY: Guilford Press.
- Boone, K. (2007). A reconsideration of the Slick et al. (1999) criteria for malingering neurocognitive dysfunction. In K. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 29-49). New York, NY: Guilford Press.
- Christiansen, H., Kis, B., Hirsch, O., Matthies, S., Hebebrand, J., Uekermann, J., . . . Philipsen, A. (2012). German validation of the Conners Adult ADHD Rating Scales (CAARS) II: Reliability, validity, diagnostic sensitivity and specificity. *European Psychiatry*, 27, 321-328.
- Conners, C. K., Erhardt, D., & Sparrow, E. (1998). *Conners' Adult ADHD Rating Scale (CAARS): ADHD across the life span*. Toronto, Ontario, Canada: Multi Health Systems.
- Dunn, T. M., Shear, P. K., Howe, S., & Ris, M. D. (2003). Detecting neuropsychological malingering: Effects of coaching and information. *Archives of Clinical Neuropsychology*, 18, 121-134.
- Erhardt, D., Epstein, J. N., Conners, C. K., Parker, J. D. A., & Sitarenios, G. (1999). Self-ratings of ADHD symptoms in adults: II. Reliability, validity, and diagnostic sensitivity. *Journal of Attention Disorders*, 3, 153-158.
- Kooij, J. J. S., Boonstra, A. M., Swinkels, S. H. N., Bekker, E. M., de Noord, I., & Buitelaar, J. K. (2008). Reliability, validity, and utility of instruments for self-report and informant report concerning symptoms of ADHD in adult patients. *Journal of Attention Disorders*, 11, 445-458.

- Lehrl, S. (1995). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B* [Multiple choice vocabulary test]. Balingen, Germany: Spitta-Verlag.
- Lehrl, S., Triebig, G., & Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurol Scand*, 91, 335-345.
- Lensing, M. B., Zeiner, P., Sandvik, L., & Opjordsmoen, S. (2013). Adults with ADHD: Use and misuse of stimulant medication as reported by patients and their primary care physicians. *Attention Deficit and Hyperactivity Disorders*, 5, 369-376.
- Murphy, K., Gordon, M., & Barkley, R. A. (2000). To what extent are ADHD symptoms common? A reanalysis of standardization data from a DSM-IV checklist. *The ADHD Report*, 8(3), 1-5.
- Musso, M. W., & Gouvier, W. D. (2014). "Why is this so hard?" A review of detection of malingered ADHD in college students. *Journal of Attention Disorders*, 18, 186-201.
- Pella, R. D., Hill, B. D., Shelton, J. T., Elliott, E., & Gouvier, W. D. (2012). Evaluation of embedded malingering indices in a non-litigating clinical sample using control, clinical, and derived groups. *Archives of Clinical Neuropsychology*, 27, 45-57.
- Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology*, 18, 379-395.
- Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical-review of methodological and clinical considerations. *Clinical Psychology Review*, 13, 255-274.
- Rösler, M., Retz-Junginger, P., Retz, W., & Stieglitz, R. (2008). *Homburger ADHS-Skalen für Erwachsene. Untersuchungsverfahren zur syndromalen und kategorialen Diagnostik der Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) im Erwachsenenalter* [Homburg ADHD scales for adults. Tools for the syndromal and categorical diagnostic of attention-deficit/hyperactivity disorder (ADHD) in adulthood]. Göttingen, Germany: Hogrefe.
- Suhr, J. A., Buelow, M., & Riddle, T. (2011). Development of an infrequency index for the CAARS. *Journal of Psychoeducational Assessment*, 29, 160-170.
- Suhr, J. A., Hammers, D., Dobbins-Buckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Archives of Clinical Neuropsychology*, 23, 521-530.
- Suhr, J. A., Zimak, E., Buelow, M., & Fox, L. (2009). Self-reported childhood attention-deficit/hyperactivity disorder symptoms are not specific to the disorder. *Comprehensive Psychiatry*, 50, 269-275.
- Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology*, 14, 189-207.
- Suslow, T. (2009). Estimating verbal intelligence in unipolar depression: Comparison of word definition and word recognition. *Nordic Journal of Psychiatry*, 63, 120-123.
- 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(Suppl. 1), S123-S134.
- Tucha, L., Sontag, T. A., Walitza, S., & Lange, K. W. (2009). Detection of malingered attention deficit hyperactivity disorder. *Attention Deficit and Hyperactivity Disorders*, 1, 47-53.
- Ward, M. F., Wender, P. H., & Reimherr, F. W. (1993). The Wender Utah rating-scale: An aid in the retrospective diagnosis of childhood attention-deficit hyperactivity disorder. *American Journal of Psychiatry*, 150, 885-890.
- Willcutt, E. G. (2012). The prevalence of DSM-IV Attention-Deficit/Hyperactivity Disorder: A meta-analytic review. *Neurotherapeutics*, 9, 490-499.