Susceptibility of functional impairment scales to noncredible responses in the clinical evaluation of adult ADHD

Anselm B. M. Fuermaier, Oliver Tucha, Janneke Koerts, Marah Butzbach, Matthias Weisbrod, Steffen Aschenbrenner & Lara Tucha


To link to this article: https://doi.org/10.1080/13854046.2017.1406143
ABSTRACT

Objective: A growing body of research questions the reliance of symptom self-reports in the clinical evaluation of attention-deficit/hyperactivity disorder (ADHD) in adulthood. A recent study suggested that also impairment reports are vulnerable to noncredible responses, as derived from a simulation design using a global functional impairment scale. The present study aims to add evidence to this issue, by using an ADHD specific impairment scale in a simulation design on large samples. Method: Impairment ratings on the Weiss Functional Impairment Rating Scale (WFIRS) of 62 patients with ADHD were compared to 142 healthy individuals who were instructed to show normal behavior. Furthermore, impairment ratings of patients with ADHD were compared to ratings of 330 healthy individuals who were randomly assigned to one of four simulation conditions that were instructed to complete the scale as if they had ADHD. Results: Patients with ADHD reported higher levels of impairment than the healthy control group in all domains of life. Furthermore, individuals instructed to feign ADHD indicated higher levels of impairments in most domains of life compared to control participants and genuine patients with ADHD. The group differences between individuals feigning ADHD and individuals with genuine ADHD, however, were only small to moderate. Further analyses revealed that the WFRIS was not useful to successfully differentiate genuine from feigned ADHD. Conclusions: The present study confirms the conclusion that self-reported impairments are susceptible to noncredible responses and should be used with caution in the clinical evaluation of adult ADHD.

ARTICLE HISTORY
Received 13 July 2017
Accepted 10 November 2017

KEYWORDS
Adult ADHD; self-report; impairment; noncredible; feigning

Susceptibility of functional impairment scales to noncredible responses in the clinical evaluation of adult ADHD

Anselm B. M. Fuermaier, Oliver Tucha, Janneke Koerts, Marah Butzbach, Matthias Weisbrod, Steffen Aschenbrenner, and Lara Tucha

Department of Clinical and Developmental Neuropsychology, University of Groningen, Groningen, The Netherlands; Department of Psychiatry and Psychotherapy, SRH Clinic Karlsbad-Langensteinbach, Karlsbad-Langensteinbach, Germany; Section for Experimental Psychopathology, Centre for Psychosocial Medicine, University of Heidelberg, Heidelberg, Germany; Department of Clinical Psychology and Neuropsychology, SRH Clinic Karlsbad-Langensteinbach, Karlsbad-Langensteinbach, Germany

Patients' self-reports represent an important and often the main source of information at diagnostic evaluation of attention-deficit/hyperactivity disorder (ADHD) in adulthood (American Psychiatric Association, 2013; Kooij et al., 2008). This is a major problem as ample

CONTACT Anselm B. M. Fuermaier a.b.m.fuermaier@rug.nl

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.
evidence exists that questions the reliability of self-reports, e.g. because a considerable number of individuals in the clinical evaluation of ADHD feign symptoms in order to get access to secondary benefits from a diagnosis of ADHD (Harrison, Edwards, & Parker, 2007; Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008). The problem of noncredible symptom report in the clinical evaluation of ADHD is emphasized by consistent findings showing that symptom self-report scales are unable to distinguish genuine patients with ADHD from individuals instructed to feign ADHD (Tucha, Fuermaier, Koerts, Groen, & Thome, 2015).

In addition to the assessment of ADHD symptom severity, a comprehensive clinical evaluation and treatment planning also requires a proper assessment of impairments in various settings of life. However, while numerous studies demonstrated that self-reported symptoms are vulnerable to noncredible responses, the vulnerability of self-reported impairments remained unknown until a very recent study of Bryant and colleagues (2017). In this study, a simulation design was performed using the World Health Organization Disability Assessment Schedule 2.0 (WHODAS; World Health Organization, 2012). Healthy individuals instructed to feign ADHD reported higher levels of impairment than healthy control participants and genuine patients with ADHD, leading the authors to conclude that the WHODAS impairment scale is susceptible to noncredible responses and should be used with caution at clinical evaluation of adult ADHD.

The present study aimed to add evidence to the susceptibility of impairment reports in the clinical evaluation of adults with ADHD. In this study, we did not use a global functional impairment scale, but instead a functional impairment scale that is widely used in clinical practice for the assessment of adults with ADHD, i.e. the Weiss Functional Impairment Rating Scale (WFIRS; Canadian ADHD Resource Alliance [CADDRA], 2017). The WFIRS has been specifically designed to measure impairment associated with ADHD in several domains and aspects of daily living, including family, work, self-concept, life skills, social functioning, and risk behavior. We conducted a large-scale simulation design with 472 healthy individuals who were assigned to either a control condition or to one of four simulation conditions and compared findings to a group of patients with ADHD.

**Method**

**Participants**

Sixty-two patients with ADHD participated in the study (Table 1). Patients were self-referred or referred from local psychiatrists or neurologists to the Department of Psychiatry and Psychotherapy of the SRH Clinic Karlsbad-Langensteinbach, Germany. A diagnostic assessment for ADHD in adulthood as well as participation in the research project was offered to all participants. It was pointed out to patients with ADHD that the clinical evaluation and

<table>
<thead>
<tr>
<th></th>
<th>CG</th>
<th>ADHD</th>
<th>SG</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>142</td>
<td>62</td>
<td>330</td>
</tr>
<tr>
<td>Age (in years, $M \pm SD$)</td>
<td>31.0 ± 12.3</td>
<td>36.2 ± 11.1</td>
<td>26.8 ± 10.0</td>
</tr>
<tr>
<td>Gender (female/male)</td>
<td>66/76</td>
<td>26/36</td>
<td>189/141</td>
</tr>
<tr>
<td>Education (in years, $M \pm SD$)</td>
<td>14.0 ± 2.6</td>
<td>11.8 ± 3.3</td>
<td>16.0 ± 2.5</td>
</tr>
</tbody>
</table>

Note: CG = Control group; ADHD = Patients with ADHD; SG = Simulation group.
the participation in the research project were separated from each other. Thus, the materials used for the research part (including the WFIRS) were not analyzed for the purpose of the clinical evaluation of individuals, but were analyzed anonymously and contributed to the research project only. Participation in the research study was separated from the clinical assessment and treatment in order to exclude possible incentives for individuals feigning ADHD to take part in the study.

Thirty-seven of the patients with ADHD met criteria for ADHD – combined type, 24 patients met criteria for ADHD – predominantly inattentive type, and one patient met criteria for ADHD – hyperactive-impulsive type, as outlined in the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM-IV; American Psychiatric Association, 1994). Twenty-five of the 62 patients with ADHD were diagnosed with one or more comorbid disorders, including mood disorders, anxiety disorder, personality disorders, and substance abuse disorders. Sixteen patients with ADHD were treated with antidepressant medication at the time of the study, whereas three patients with ADHD were currently treated with stimulant medication.

Furthermore, 472 healthy individuals participated in the study (Table 1). Healthy individuals were recruited via public announcements, word-of-mouth and through contacts of the researchers involved. None of the healthy individuals reported to have a history of neurological or psychiatric disease and none were taking any medication known to affect the central nervous system. Furthermore, all healthy participants scored below the clinical cutoffs of two standardized self-report rating scales for current and retrospective ADHD symptoms (Adler et al., 2006; Kessler et al., 2005; Ward, Wender, & Reimherr, 1993).

**Materials**

The Weiss Functional Impairment Rating Scale (WFIRS) is a self-report measure for impairments that commonly occur in patients with ADHD and that are likely to represent the patients’ targets of treatment (CADDRA, 2017). The WFIRS comprises 70 items that are divided into seven domains: *Family* (8 items), *Work* (11 items), *School* (11 items), *Life Skills* (12 items), *Self-concept* (5 items), *Social* (9 items), and *Risk* (14 items). The domain *School* was not considered in the present study as the majority of individuals did not follow an educational training at the time of participation. Each item employs a four-point Likert scale scored from 0 to 3 (*0 = never, not at all; 1 = sometimes, somewhat; 2 = often, much; 3 = very often, very much*). An additional option to answer is given with *Not Applicable*. A scale score per domain is calculated by summing up the responses to all items per domain (response values 0 through 3), and dividing this sum by the number of endorsed items (thereby not considering items that are answered with *Not Applicable*). The WFIRS was reported to yield high internal consistency with Cronbach’s α > .8 for each domain and the scale as a whole. Furthermore, the WFIRS was reported to be sensitive to change with treatment and to correlate with change in ADHD symptoms and overall psychopathology (CADDRA, 2017).

In addition, three measures of performance validity were applied in order to ensure the credibility of patients with ADHD, i.e. the Test of Memory Malingering (TOMM; Tombaugh, 1997), the Dot Counting Test (DCT; Boone, Lu, & Herzberg, 2002), and the Groningen Effort Test (GET; Fuermaier, Tucha, Koerts, Aschenbrenner, & Tucha, 2016; Fuermaier, Tucha, Koerts, Grabski, et al., 2016). Cut-offs were applied as suggested in the test manuals.
Design and procedure

Participation in the research study was separated from diagnostic assessment of patients with ADHD. Diagnostic assessment for ADHD involved a clinical psychiatric interview according to DSM-IV criteria and was supported by two standardized self-report rating scales designed to quantify current and retrospective ADHD symptoms (Adler et al., 2006; Kessler et al., 2005; Ward et al., 1993). Diagnostic veracity was corroborated by the identification of objective evidence of impairment (e.g. financial problems, failure in academic setting, losing jobs, drug use) and/or the consult of collateral information. The research study was performed in the years from 2013 to 2016 and was conducted in compliance with ethical standards and was approved by local institutional ethical committees. Patients were only considered for inclusion in the study if they passed all effort tests that were presented to them. The completion of the WFIRS for patients with ADHD was part of a larger assessment which took about 2.5 h in total. Seventy-four patients with ADHD were initially considered for inclusion in the present study. Forty-eight patients with ADHD performed the TOMM and DCT, whereas the GET was presented to 26 patients with ADHD. Twelve patients with ADHD (16%) were not further considered as they failed the effort measures presented to them. The remaining 62 patients with ADHD were asked to fill out the WFIRS to the best of their knowledge and not to seek help from the examiner or to discuss questions or their responses.

The assessment of healthy individuals took about 30–60 min, depending on the condition they were assigned to. All healthy participants completed a demographic questionnaire, two self-report rating scales for current and retrospective ADHD symptoms, as well as the WFIRS. Instruction how to complete the WFIRS differed between the conditions healthy individuals were assigned to. Two control groups were recruited; control group 1 (n = 67) was recruited in order to obtain a matched group to patients with ADHD with comparable characteristics in age, gender, and years of education. Control group 2 (n = 75) was obtained from the simulation design in which 405 healthy participants were randomly assigned to either the control group 2 (n = 75), or to one of four simulation conditions that varied by type and level of coaching. Participants of the control groups were asked to complete the WFIRS to the best of their knowledge. WFIRS scores of both control groups did not differ significantly, therefore control groups 1 and 2 were collapsed into one single control group (CG; n = 142) for further analyses. Participants of the simulation conditions were asked to complete the WFIRS while pretending to be affected by ADHD. Because simulation group differences were not the focus of this study, these four conditions were combined into one simulation group (SG: n = 330).

Statistical analysis

Because assumptions for parametric analyses were not met (e.g. normality, homogeneity of variances), nonparametric statistical analyses (Kruskal–Wallis tests) were calculated to compare the groups on each domain of the WFIRS. Dunn’s tests were performed as post hoc multiple pairwise comparisons. Significance level was initially set to .05, but was Bonferroni corrected in order to control for alpha error growth in multiple testing (i.e. .05/6 = .008 to control for pairwise comparisons on each of the six domains of the WFIRS). Effect sizes (Cohen’s d) were calculated to indicate the magnitude of pairwise group differences. Effect sizes were interpreted based on Rogers’ classification into moderate effects (.75 ≤ d < 1.25),
large effects (1.25 ≤ d < 1.50), and very large effects (d ≥ 1.50) (Rogers, 2008). Furthermore, for those scales of the WFIRS that revealed significant differences between patients with ADHD and the simulation group, receiver operating characteristics (ROC) were calculated in order to explore the utility of the WFIRS in distinguishing patients with ADHD from individuals instructed to feign ADHD. Clinical utility was also examined by classification statistics of sensitivity, specificity, positive predictive value (ppv), and negative predictive value (npv).

**Results**

WFIRS scores of healthy control participants, patients with ADHD, and the simulation group are presented in Table 2. Nonparametric statistical analyses as presented in Table 2 revealed significant differences between the three groups in each domain of the WFIRS. Post-hoc multiple pairwise comparisons indicated that both patients with ADHD and the simulation group scored significantly higher on each scale of the WFIRS than the control group. Furthermore, the simulation group scored significantly higher than patients with ADHD in the domains *Family*, *Work*, *Social*, and *Risk*, whereas no significant differences were obtained in *Life Skills* and *Self-concept*. Effect sizes of group differences between patients with ADHD and the simulation group were negligible to moderate.

Finally, ROC analyses demonstrated that the scales *Family*, *Work*, *Social*, and *Risk* could distinguish significantly between feigned ADHD and genuine ADHD (all p-values ≤ .002); however, classification accuracies were only moderate as shown by area under the curves (AUCs) of 64.3% (*Family*), 71.4% (*Work*), 62.9% (*Social*), and 78.1% (*Risk*). A graphical illustration of diagnostic accuracies of the four scales of the WFIRS is presented in Figure 1. The limited clinical utility of the WFIRS in distinguishing genuine from feigned ADHD was also shown by classification statistics of sensitivity, specificity, ppv, and npv. Setting the cut-off on each scale to achieve desired specificity of 90% (Boone, 2007; Marshall et al., 2010) resulted in insufficient or only moderate classification accuracies of all scales (Table 3).

**Discussion**

As expected, patients with ADHD in the present study reported significantly higher levels of impairment than healthy control individuals in all domains of life. More importantly for the present context, instructed simulators reported significantly higher levels of impairment than control participants in each of the six domains, and higher levels of impairment than genuine patients with ADHD in four domains of life. However, effect sizes between genuine patients with ADHD and instructed simulators were only small to moderate, which implies that one might not be able to distinguish genuine from feigned ADHD based on individual impairment reports. This is also illustrated by ROC analyses which revealed that WFIRS scores were significantly predictive for feigned ADHD relative to genuine ADHD; however, diagnostic accuracies were only moderate (AUCs ranging from 62.9 to 78.1%) and question the clinical utility of the WFIRS for the detection of feigned ADHD.

The results of the present study confirm previous research showing that self-reported impairment scales are susceptible to noncredible responses in the clinical evaluation of adult ADHD and agree with the statement of Bryant and colleagues that ‘the assessment for the validity of self-report and performance should be included in all evaluations, including assessments for ADHD’ (Bryant et al., 2017; Heilbronner et al., 2009). The use of measures of
Table 2. Comparison of Weiss Functional Impairment Rating Scale (WFIRS) scores between healthy control participants (CG), patients with ADHD (ADHD), and the simulation group (SG).

<table>
<thead>
<tr>
<th>Domain of impairment</th>
<th>CG (n = 142)</th>
<th>ADHD (n = 62)</th>
<th>SG (n = 330)</th>
<th>Omnibus test&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Pairwise group comparison&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Pairwise group comparison&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Pairwise group comparison&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M ± SD</td>
<td>M ± SD</td>
<td>M ± SD</td>
<td>X² (df), p</td>
<td>Z, p, Cohen's d</td>
<td>Z, p, Cohen's d</td>
<td>Z, p, Cohen's d</td>
</tr>
<tr>
<td>Family</td>
<td>.57 ± .49</td>
<td>1.37 ± .63</td>
<td>1.67 ± .57</td>
<td>223.7 (2), &lt;.001</td>
<td>6.89, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.49, 14.66, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>2.01, 3.27, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>Work</td>
<td>.46 ± .49</td>
<td>1.28 ± .71</td>
<td>1.75 ± .54</td>
<td>248.3 (2), &lt;.001</td>
<td>6.14, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.45, 15.72, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>2.46, 4.37, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>Life skills</td>
<td>.67 ± .46</td>
<td>1.36 ± .62</td>
<td>1.55 ± .51</td>
<td>194.8 (2), &lt;.001</td>
<td>7.08, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.34, 13.94, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.78, 2.32, .020</td>
</tr>
<tr>
<td>Self-concept</td>
<td>.66 ± .60</td>
<td>1.87 ± .86</td>
<td>1.65 ± .68</td>
<td>166.4 (2), &lt;.001</td>
<td>12.10, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.76, 9.55, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.51, 1.73, .084</td>
</tr>
<tr>
<td>Social</td>
<td>.54 ± .47</td>
<td>1.26 ± .68</td>
<td>1.58 ± .55</td>
<td>215.4 (2), &lt;.001</td>
<td>6.76, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.33, 14.68, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.97, 3.20, .001</td>
</tr>
<tr>
<td>Risk</td>
<td>.52 ± .45</td>
<td>.78 ± .52</td>
<td>1.34 ± .57</td>
<td>192.5 (2), &lt;.001</td>
<td>2.97, .003&lt;sup&gt;+&lt;/sup&gt;</td>
<td>.55, 13.38, &lt;.001&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.53, 6.48, &lt;.001</td>
</tr>
</tbody>
</table>

Notes: CG = Control group; ADHD = Patients with ADHD; SG = Simulation group;
<sup>a</sup>Kruskal–Wallis Test.
<sup>b</sup>Dunn's post hoc test.
<sup>+</sup>Statistical significant at Bonferroni corrected significance level of .05/6 = .008.
both performance and symptom validity appears particularly relevant given the findings of only weak associations between both types of measures in the clinical evaluation of adults with ADHD (Hirsch & Christiansen, 2015). With regard to patients’ self-reports, previous studies as well as the present data revealed an over-reporting response bias of noncredible individuals relative to genuine patients with ADHD (Fuermaier, Tucha, Koerts, Weisbrod, et al., 2016; Harrison & Armstrong, 2016; Suhr, Buelow, & Riddle, 2011). Clinicians aiming to assess symptom validity are therefore advised to make use of scales consisting of items that are infrequently endorsed by healthy individuals and genuine patients with ADHD, in order to successfully distinguish credible from noncredible individuals. While such scales are not ready for use in clinical practice, yet, promising work has been done in the development and validation of infrequency scales or exaggeration indices that consist of items that are embedded within valid and widely used self-report inventories, as this would make the detection strategy less obvious to individuals feigning the condition (Cook, Bolinger, & Suhr, 2016; Harrison & Armstrong, 2016; Suhr et al., 2011).

Figure 1. Receiver operating characteristic (ROC) curves indicating diagnostic accuracies of four scales of the Weiss Functional Impairment Rating Scale (WFIRS) in identifying feigned ADHD (n = 330) relative to genuine ADHD (n = 62).

Table 3. Classification accuracies of four scales of the Weiss Functional Impairment Rating Scale (WFIRS) in identifying feigned ADHD (n = 330) relative to genuine ADHD (n = 62).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Cutoff</th>
<th>Specificity (%)</th>
<th>Sensitivity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>2.14</td>
<td>89.7</td>
<td>17.1</td>
<td>88.9</td>
<td>16.8</td>
</tr>
<tr>
<td>Work</td>
<td>2.29</td>
<td>89.7</td>
<td>13.5</td>
<td>88.0</td>
<td>19.3</td>
</tr>
<tr>
<td>Social</td>
<td>2.12</td>
<td>89.7</td>
<td>13.8</td>
<td>88.2</td>
<td>16.4</td>
</tr>
<tr>
<td>Risk</td>
<td>1.42</td>
<td>89.7</td>
<td>47.7</td>
<td>95.7</td>
<td>24.1</td>
</tr>
</tbody>
</table>

Note: PPV = Positive predictive value; NPV = Negative predictive value.
Of note, effect sizes of group differences between patients with ADHD and instructed simulators in the present study were smaller compared to most of the effects reported by Bryant and colleagues (2017). This could be explained by the use of an ADHD-specific scale in the present study. It can be assumed that a disease-specific scale likely results in larger impairment scores of adults with ADHD, which may reduce the over-reporting effect of individuals instructed to feign ADHD relative to genuine ADHD patients. The use of an ADHD-specific impairment scale could also account for the lack of significant group differences between patients with ADHD and instructed simulators on Self-concept, as patients with ADHD indicated marked impairments in this domain. The group of patients with ADHD had a mean score of 1.87 in this domain, while any score >1.5 is indicative for impairment (CADDRA, 2017). Furthermore, in contrast to the sample of patients included in the study of Bryant and colleagues (2017), the vast majority (about 95%) of patients with ADHD of the present study was not treated with stimulant medication. The early stage in the treatment process may have resulted in the endorsement of higher levels of impairment of patients with ADHD, which may have also reduced the over-reporting effect of individuals instructed to feign ADHD.

It must also be noted that the present study did not focus on student samples but included participants from the local community with regard to both the patient group and the experimental groups of the simulation design. The present study therefore revealed that the findings previously reported on student samples can be generalized to the non-academic setting.

**Limitations**

The present study must be seen in the context of some limitations. First, patients with ADHD of the present study were screened with one or two performance validity measures, and were included in the study only if they passed the measures presented to them. This may be problematic because this study focused on the credibility of self-reports of patients with ADHD, however, previous literature has shown that performance validity does not always translate into symptom validity (Hirsch & Christiansen, 2015). The inclusion of patients with ADHD that have been screened for both performance as well as symptom validity would strengthen the validity of the conclusions. In this context, it would also be interesting to include another clinical comparison group to the present study design, i.e. a group of patients with ADHD that failed measures of symptom and performance validity.

Furthermore, simulation designs in general can be criticized for a limited external validity, as the motive to feign ADHD in an experimental setting does not match real-life situations in clinical practice (Rogers, Harrell, & Liff, 1993). The validity of conclusions drawn from studies employing simulation designs would therefore benefit from studies using different research designs, such as known-groups comparisons.

**Conclusion**

The present study demonstrated that a widely used self-report scale assessing functional impairment of adults with ADHD is susceptible to noncredible self-reports. Clinicians should be prepared that over-reporting of symptoms and impairments may occur frequently in
clinical practice, and that the inclusion of proven measures assessing symptom and performance validity in the clinical evaluation of adults with ADHD appears necessary.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**References**


