rijksuniversiteit groningen

The inconsistency strategy as detection method for malingering in assessment of adult attention-deficit/hyperactivity-disorder

Elisabeth Maria van der Moolen

Master Thesis – Clinical Neuropsychology

S*3959333*
March 2024
Department of Psychology
University of Groningen
Examiner/Daily supervisor: Dr.
Anselm B.M. Fuermaier

Abstract

This simulation study investigated the utility of performance variability scores as a strategy for the detection of feigned adult attention-deficit/hyperactivity disorder (ADHD). Three groups of participants completed three performance validity tests on two separate occasions, with an average of 32 days in between assessments. The patient group consisted of individuals with an ADHD diagnosis who were asked to perform to their honest abilities, the control group consisted of healthy individuals asked to perform to their honest abilities, and the simulation group consisted of healthy individuals who were asked to feign ADHD as convincingly as possible. Variability scores were derived from the test scores of the two assessments. The variability scores were not helpful in the detection of feigned ADHD when used as only method and did not add value to a model in which single test scores were already incorporated. The consistency method does not seem a promising research venture for the detection of feigned ADHD in clinical assessments.

Key words: ADHD, malingering, performance validity, consistency

The inconsistency strategy as detection method for malingering in assessment of adult attention-deficit/hyperactivity-disorder

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that is characterized by inattention, hyperactivity and impulsivity, and the symptoms are persistent and impairing in daily life (Wallace et al., 2019).

The psychiatric diagnosis of ADHD can be a challenging process (Braun et al., 2004; Greenfield et al., 2002). Because there are no physical markers or somatic tests that can indicate ADHD in a person, the diagnosis is focused on the behavioral difficulties and cognitive deficits that people with ADHD present. However, though cognitive problems are part of ADHD and the diagnostic process, no single cognitive deficit on its own is indicative of ADHD. The symptoms of ADHD are all 'normal' human behaviors that are experienced by most individuals at some point in their lives (Murphy & Adler, 2004), and the dimensional characteristic of the disorder makes it difficult to decide cutoffs for when it is still regular behavior and when it becomes a disorder (Faraone et al., 2000). Furthermore, these problems must be persistent over time and are seen less clear and direct in assessment with a clinician than in the daily life context of those who experience them in school or at work (Marshall et al., 2016).

Besides ADHD being a difficult diagnosis to make in general, it is more difficult to diagnose in adults than in children or adolescents. A reason for this difficulty is that for an ADHD diagnosis according to DSM criteria, several symptoms must be present before the age of 12 years. This means that part of the adult ADHD diagnosis will be evaluated retrospectively. However, a major difficulty in retrospective diagnosis is that it tends to be biased by current functioning (McGough & Barkley, 2004). This bias comes from adults, especially highly intelligent individuals, often being able to compensate for difficulties they

experienced from their undiagnosed ADHD (Primich & Lennaco, 2012). It is only when daily-life demands exceed their abilities to compensate, that the impairments show up clearly. This is why adult patients are often unable to accurately identify symptoms from an early age as they did not know they were compensating. This also affirms the need for information from life-contexts such as work or school, but those are not always accurate or even available. Until developmentally appropriate symptoms and diagnostic thresholds for adult ADHD are adequately defined, clinicians must practice clinical judgment when applying the DSM criteria for an ADHD diagnosis to adult patients. Without well-validated and universally accepted diagnostic criteria, the risk of overdiagnosis and underdiagnosis of adult ADHD remains a challenge for clinicians (McGough & Barkley, 2004).

A common phenomenon in clinical assessment in general and in the clinical assessment of ADHD is noncredible or invalid symptom reporting and performance on assessment tests. This phenomenon presents a challenge on its own for clinicians as this may cause biased or false diagnoses and complicates assessment. Furthermore, this then causes problems with finding appropriate treatments (Hirsch et al., 2022). Underlying reasons for noncredible symptom reporting or performances on tests can be difficult to determine, as it can be either conscious or unconscious, depend on the sample and context of the assessment and varying motivations may apply (Dong et al., 2023). One possible underlying reason may be the deliberate feigning of ADHD motivated by external incentives, also referred to as malingering. Base rates of noncredible symptom report or performance depend on the population studied and context of the assessment (Hirsch et al., 2022). Differences in research methodology, criteria to determine noncredible performance and different sample compositions across studies cause varying base rate estimations. In college students evaluated on their own initiative for example, different studies estimated noncredible symptom report

and -performance to occur from 12 – 48% of assessments (Harrison et al., 2010; Sullivan et al., 2007).

There is a variety of external incentives that can drive a person to feign symptoms, such as medical, legal or financial gains like prescription drugs, milder legal punishment or financial aid. As aforementioned, exact prevalence of feigning is still unknown, but several researchers agree that it is likely for rates to be higher in forensic settings than in clinical settings because in forensic settings punishment avoidance is also a contributing factor (Bianchini et al., 2005; Greve et al., 2013; Mittenberg et al., 2002). In the clinical context, feigning may be especially high among university students due to the specific academic benefits it may bring them, such as stimulant medication, or extra time on exams or a quiet test-taking environment (Harp et al., 2011). Another reason for people to feign ADHD may be because they experience numerous difficulties in life that they may associate with their self-esteem and abilities and would like an alternative explanation for. This way, it may make them feel less responsible in terms of personal effort (Wei & Suhr, 2015). They may do this consciously, but this can also happen unconsciously. In research of Dandachi-FitzGerald et al. (2020) about feigning in the everyday context, the authors even argue that neuropsychological literature on malingering has been one-sided as it has largely emphasized external incentives as motivation. In their study they found that psychological motives such as excuse making and seeking attention from others seem to be notably important in everyday feigning.

Malingering in the educational setting is concerning because university students were found to be especially successful at it. Various studies have shown that students can convincingly and easily modify their answers on ADHD behavioral rating scales in ways that make it look like they have ADHD (Conners et al., 1999; Dupaul et al., 1998; Fisher & Watkins, 2008; Harp et al., 2011; Harrison et al., 2007; Jachimowicz & Geiselman, 2004; Jasinksi et al, 2011; Rios & Morey, 2013; Tucha et al., 2009).

Malingering in ADHD assessments is detrimental to society as negative consequences are manifold, such as the costs of assessments and treatments that are unjustified, the medical resources that are being used and the consequences of unnecessary and incorrect drug use. Malingering also creates more stigma around therapies for ADHD and reduces societal confidence in therapy. Those who truly have ADHD are also affected by these consequences and for them this issue may cause the greatest disadvantage. This is because the measures taken to, for example, improve medication control are added barriers for patients to easily access their needed treatment (Fuermaier et al., 2016). These consequences stress the need for accurate detection methods for malingering during ADHD assessments.

Previous literature shows that detection accuracy of feigned ADHD varies. This is partly explained by the variation in ADHD assessment methods, as the detection methods need to be accustomed to each type of evaluation. Studies have shown that individuals who are instructed to feign ADHD on self-report symptom scales successfully produce results that are accurate for the disorder (Harrison et al., 2007; Quinn, 2003; Williamson et al., 2014). Even though research is being done on this topic, not many self-report measures or neuropsychological tests have shown to be very accurate and sensitive in differentiating feigned ADHD from real ADHD (Musso & Gouvier, 2014). A method of malingering detection that focuses on symptom validity is called a symptom validity test (SVT). These tests can be either stand-alone, as a separate test, or embedded in an existing symptom report questionnaire. An example of an embedded SVT is the validity test in the Conners' Adult ADHD Rating Scale (CAARS), which is called the CAARS Infrequency Index (Suhr et al., 2011). Embedded tests are especially practical for clinicians as they do not take up extra time during evaluations.

Another method of malingering detection focuses on validity of performance on cognitive tests during ADHD assessment. This kind of test is called the performance validity

test (PVT). PVTs can also be either stand-alone or embedded. An example of a stand-alone PVT is the Test of Memory Malingering (TOMM; Tombaugh, 1996). This is a visual recognition test that consists of trials for learning and one trial for retention. These types of tests are specifically designed to detect invalid performance. Another stand-alone attention-based PVT that was developed more recently is the Groningen Effort Test (GET; Fuermaier et al., 2017; Fuermaier et al., 2016; Fuermaier et al., 2020). This test appears to the participant to be an attention test but is a visuospatial perceptual task. For adult ADHD the GET is very accurate in detecting feigned attention deficits, which was shown in a simulation design of Fuermaier et al. (2020) where the GET showed 88% sensitivity for feigned ADHD and 90% specificity for genuine ADHD. In a study with a simulation design healthy participants are randomly allocated to different conditions, in which they are asked to complete rating scales and tests in particular ways. Participants of the control condition are asked to complete tests and rating scales honestly and with genuine effort while participants of the experimental condition receive instructions to feign symptoms or diseases. The responses of these instructed participants are usually compared to the responses of participants in the control group and genuine patients of the disorder or disease targeted in the study (Tucha et al., 2015).

An example of an embedded indicator in a PVT is the Reliable Digit Span (RDS) from the Digit Span Test of the Wechsler Adult Intelligence Scale (WAIS-IV; Petermann, 2012). This is a calculated score that acts as predictive measure for noncredible responding. Previous studies have shown that the RDS can detect ingenuine effort in various clinical adult populations, such as adult ADHD (Rogers et al., 2021).

Another test that can be used as an embedded PVT is the Continuous Performance Test (CPT). This test was designed to assess one or several aspects of sustained attention and vigilance. The task characteristics of CPTs vary widely and stimulus materials are different

across tests. The test variables that are derived from CPTs, to serve as predictors for noncredible cognitive performance, are most often response style, expressed through errors, and response time, expressed through reaction time and its' variability (Dong et al., 2023). Currently, the most widely used CPT is the Conners' Continuous Performance Test (C-CPT). The most recent version, the third edition (C-CPT-3; Conners, 2014) measures sustained attention through measuring inattention, impulsivity, and vigilance. Another example of a CPT comes from the Vienna Testing System (VTS; Schuhfried, 2013) and is called the Perception and Attention Functions - vigilance (WAFV; Fuermaier et al., 2022).

While cognitive testing is not required in evaluations for an ADHD diagnosis (American Psychological Association, 2013), it may bring benefits with regards to malingering detection. PVTs are thought to be less vulnerable for coaching, which means that it is harder for individuals wanting to feign ADHD to successfully prepare themselves on a performance test (Quinn, 2003). This way, performance may be determined to be statistically significantly below the performance level that most genuine individuals with ADHD could score, or by chance levels (Slick et al., 1999). With this should be mentioned that a downside of using PVTs is that it would lengthen assessments for the clinician. Nonetheless, considering the alarming consequences of malingering in ADHD assessments, and as both SVTs and PVTs are still being developed and improved to better and more consistently detect malingering, Suhr and Berry (2017) argue they both should be included in clinical assessment.

Besides the infrequently endorsed symptoms and symptom exaggeration some clinicians claim a highly variable (inconsistent) symptom presentation to be a warning sign of feigning as well. In previous literature this has not been extensively researched yet (Bianchini et al., 2005; McDermott & Feldman, 2007). The medical model states that symptoms are signals of underlying problems which are stable and remain permanent unless they are treated (Blaney, 2014) and from this model stems the idea that symptom variability within

individuals is a red flag for feigning. This is similar to the general public's beliefs about deception (Granhag & Stromwall, 1999). They have the assumption that consistent symptom reports point to honest responding and inconsistencies are a signal for deception. This also shows the assumption that feigners of symptoms are unable to convincingly do so for longer periods of time (Jelicic et al., 2017). However, cases in which people were convincingly feigning symptoms for weeks have been reported (van der Heide et al., 2020). In specific conditions, such as trauma, studies have even shown that feigned symptoms were more consistent over longer periods of time than genuine symptoms, so whether this holds true or not remains debated (Peace et al., 2010) and it is still unclear how this would show up in ADHD symptomology.

In the existing body of literature on malingering detection in ADHD assessment the inconsistency strategy has been researched for SVTs, like in a study of Boskovic et al. (2022). In this study the authors investigated how individuals experiencing genuine pain-symptoms and individuals feigning such symptoms rated the intensity of their symptoms over a period of five days. With this it was also investigated if these ratings showed intraindividual (in)consistency. Exaggeration of symptom intensity resulted to be distinguishing between the two groups, but inconsistency was not found significant. However, the authors noted that a limitation of their study was that five days may not have been a long enough period for intraindividual response fluctuations to be captured. Therefore, investigating this for longer periods of time may yield different results.

Furthermore, regarding the consistency strategy and PVTs, it is known that in research of Robinson et al. (2023) repeated administration of the Conners Continuous Performance Test-II (CPT-II) improved its classification accuracy as a performance validity index. In this study the CPT-II was administered to a sample of mostly psychiatric patients, and a small proportion of patients with mild traumatic brain injury. The CPT-II was administered in the

morning and again at the end of the testing appointment. Results showed that while a certain level of uncertainty about performance validity exists with a single administration, this can partly be addressed with a second administration. As the CPT-II is computer-based, it is automatically scored and self-administered, which would make routine application of repeated administration in clinical assessments more feasible. This is promising for the concept of repeated administration of validity testing instruments in ADHD assessment with regards to malingering detection.

Based on the current state of literature the potential of the consistency strategy in the context of ADHD malingering detection with longer periods of time between PVT administrations has not been adequately addressed. The present study employed a simulation design in which three PVTs were administered twice to a mixed sample of 61 individuals. A total of 8 variables were drawn from results of the GET, Digit Span Test and WAFV as measures of performance validity and difference scores were calculated to measure (in)consistency of performance. The three groups in this study were a patient-group of individuals diagnosed with ADHD, a simulation-group of individuals who received instructions to feign ADHD in the experiment and a control-group of individuals who were instructed to respond honestly and to the best of their ability in the experiment. The absolute variability scores of these groups were compared to investigate their level of (in)consistency in performance. The aim of this study was to investigate whether feigned ADHD could be distinguished from genuine ADHD and honest responding individuals using variability (inconsistency) in performance on repeated administration of PVTs over longer periods of time. Based on prior research and the aim of this study the following research hypotheses were formulated: (1) The control-group may show the smallest absolute variability scores and thus show the most consistent performance, (2) The patient-group may show larger absolute variability scores than the control-group, but smaller absolute variability scores than the

simulation-group, thus performing less consistent than the control-group but more consistent than the simulation-group, (3) The simulation group may show the largest absolute variability scores of all the groups and thus show the least consistent performance, and (4) The differences between the groups may be larger for the variability scores than the differences between the groups at the first assessment (the first assessment is denoted as time-point one; T1), thus showing the incremental value of repeated assessment and variability scores.

Methods

Participants

Patient group

The 18 patients with ADHD that participated in the study were recruited from and assessed at the outpatient unit of the department of psychiatry and psychotherapy of SRH Hospital Karlsbad-Langensteinbach, Germany. Demographic information of these patients is presented in table 1. These patients all had an established ADHD diagnosis. They were provided a financial reward for participation. Current psychiatric diagnoses that were present in some participants in the patient group during the study were psychiatric- and behavioral disorder from use of substances, mood disorders, affective disorders and adjustment disorder.

Simulation and control groups

All 27 participants in the simulation group and 18 participants of the control group were recruited through a convenience sample. Demographic information of these participants is presented in table 1. Allocation of participants to either simulation group or control group was randomized. No specific selection procedure was implemented, but criteria were adhered such as the exclusion criterium that participants for these groups could not have an established ADHD diagnosis, and inclusion criterium that participant age had to be approximately between 20 and 60 years and they had to have an adequate comprehension of the English

language. Participants were approached and provided with sufficient information about the study to be able to consider participation. No compensation was provided for participation, as it was on a voluntary basis. Any psychiatric diagnoses that were present in some participants in the simulation-group during the study were dyslexia and personality disorder.

Table 1

Demographic information per group

Characteristic:	Simulation group	Control group	Patient group
Total N	27	16	18
Sex (0/1)	9/18	8/8	6/12
Education level (1/2/3/4/5)	0/4/3/11/9	0/0/2/8/6	0/1/7/6/4
Mean age (SD) (minimum; maximum)	22 (2)(19;32)	25(4)(21;37)	45(11)(20;65)
Mean time between assessments (minimum; maximum)	34 (19;55)	35(20;66)	30(20;48)

Note. Education levels: 1 = none; 2 = compulsory schooling or intermediate secondary school; 3 = college or vocational training; 4 = higher secondary school with university entrance qualification; 5 = university degree. Sex: 0 = female; 1 = male. Mean age is expressed in years. Mean time between assessments is expressed in days.

Materials and measures

Three PVTs were administered, one stand-alone test and two embedded tests. The stand-alone test that was administered was the Groningen Effort Test (GET; Fuermaier et al., 2016; Fuermaier et al., 2017; Fuermaier et al., 2020). This is a computer-based test which participants completed using a laptop with the Vienna Testing System keyboard attached. The GET is a test that to the participant appears to be an effort test but is a visuospatial perceptual test that measures the ability to suppress central coherence (Fuermaier et al., 2016, 2017). In each trial, participants are shown a target shape and a complex figure. The participant must respond whether they think the target shape is present in the complex figure or not, by pressing buttons on the keyboard. From the GET the measures GET index (GETINDEX), total errors (GETERR), and mean processing time (GETRT) were derived.

The first embedded test that was administered was the Digit Span Test, from the WAIS-IV (Petermann, 2012). This test is administered and completed verbally while the administrator takes note of the participant's responses. The Digit Span Test consists of a forward and a backward version in which the administrator reads a series of numbers aloud and the participant must recall them in the correct order or, in case of the backward version, in the opposite order. The series of numbers become longer as the test continues. The test is discontinued when a participant either finishes all the trials, or when they fail two series of numbers in the same trial. Both the forward and backward version were administered and from those the measure Reliable Digit Span (RDS) was calculated. The RDS is derived from adding up the longest digit span forward and the longest digit span backward in which both trials were passed (Greiffenstein et al., 1994).

The second embedded test that was administered was the WAFV (WAFV; Fuermaier et al., 2022). This is a computer-based test which participants completed using a laptop with

the Vienna Testing System Keyboard attached. The long version of this test is a visual test in which 900 squares are presented on a screen, one at a time, and participants are required to press a response button as quickly as possible when the square becomes darker. These darker squares form the targets, which form a frequency of 5% of the total number of presented stimuli. This test takes 30 minutes to complete. The measures that were derived from this test include the number of omission errors (WAFVMISS), commission errors (WAFVFALSE), the mean response time of correct responses (milliseconds) (WAFVRT), and the standard deviation of the response time (WAFVRTSD; Fuermaier et al., 2022).

Procedure

This study was part of a larger research project. The procedure relevant to the current study is described.

Patient group

Approval of the assessment of patients with ADHD was given by the medical ethical committee of University of Heidelberg, Germany. All evaluations were completed between January 2019 and March 2020. The part of day at which the evaluations took place was attempted to be generally the same for both the first and the second assessment. The average time between the two administrations was 30 days. The evaluations took place at the outpatient unit of the department of psychiatry and psychotherapy of SRH Hospital Karlsbad-Langensteinbach, Germany. The patients with ADHD were first presented with information about the study, and an informed consent form. All patients that participated in the study signed the informed consent. Demographic information was then collected through a questionnaire that included questions about their age, gender, highest education level completed, other current psychiatric conditions, and if they were taking medication that affects the nervous system. Participants then completed the Digit Span, the GET computer task, and the WAFV computer task. After completion of the tasks the participant was

debriefed. The researcher was present the entire duration of every evaluation, to administer the tests but also to answer any questions participants might have had during the experiment. The researcher did not engage in distracting behavior or attempts of participants to make small talk during completion of the tests. In case of any technological problems the researcher rectified those when possible.

Simulation and control groups

The procedure was approved by the Ethical Review Board of the Faculty of Behavioral- and Social Sciences, University of Groningen, Netherlands, with an approval date of 31 October 2022, approval number PSY-2223-S-0029. The experimental setup that was chosen was a simulation design in which all evaluations were completed between November 2022 and March 2023. The part of day at which the evaluations took place was attempted to be generally the same for both the first and the second assessment but this differed per participant. The average time between the two administrations was 32 days. The evaluations took place either at the researcher's house or at the participant's house.

The evaluations started with the researcher introducing the experiment and giving information about which activities the evaluation would include, such as the filling out of some questionnaires and the completion of attention tests. An informed consent form was provided to all participants after the information right at the beginning of the study. All participants who entered the study signed informed consent and were informed that withdrawal would be accepted at any time during the evaluations without any explanation needed. Participants then completed questions about their demographics, which included their age, gender, highest education level that they completed, whether there was presence of an ADHD diagnosis in their childhood or adulthood, whether they had any other psychiatric conditions, and whether they were currently taking any medications affecting the nervous system. Depending on the condition of the participant, instructions were given on how to

complete the tasks in the experiment. The control-group received instructions to complete the three consecutive tests honestly and to the best of their ability. The simulation-group received information about the symptoms of ADHD as described in the DSM-5, with criteria to fulfill regarding both hyperactivity and inattention, and then were asked to perform the following tests as if they had ADHD. They were asked to do so as convincingly as they could. Before they started with the first task a questionnaire was filled out, as pre-simulation check, to control if the participant really understood the instructions they had just received. Questions were asked such as "Which symptoms belong to ADHD in adulthood?", and "How are you supposed to complete the following tasks?". They could choose from multiple-choice answers such as "symptoms of inattention, hyperactivity, and visual disorders", "symptoms of inattention, hyperactivity, and impulsivity", and "symptoms of hyper-arousal, hyperactivity and hypertension" for the first question. Here, the second option would show that the participant adequately understood the information they received about the symptoms of adult ADHD. Regarding the second question, participants could choose from the multiple-choice answers of "as if I had ADHD", "Honestly", and "I will see". Here, the first option would be indicative of adequate understanding of the participant about the instructions to convincingly feign ADHD during the upcoming tasks. Then they proceeded to complete the tasks, the Digit Span, the GET computer task, and the WAFV computer task consecutively. After completion of the tests, participants were informed they could stop feigning ADHD. A post-experimental check questionnaire was then completed to check once again if they still understood the assignment after the testing. This questionnaire included questions such as "Did you complete the previous three tasks honestly?" answered by either 'yes' or 'no', "Did you adhere to the instructions and faked ADHD?" answered by either 'yes' or 'no', and if answered 'yes' to the question whether they adhered to the ADHD feigning instructions, they were asked to answer on a scale of one to five how much they tried to fake ADHD and how successful they thought

they were in faking ADHD, one being ‘hardly’ and five being ‘very much’. This served as indicator if they continued to apply the correct instructions during the testing. After the post-experimental check, the participant was debriefed.

The researcher was present the entire duration of every evaluation, to administer the tests but also to answer any questions participants might have had during the experiment. The researcher did not engage in distracting behavior or attempts of participants to make small talk during completion of the tests. In case of any technological problems the researcher rectified those when possible.

Statistical Analysis

Basic descriptive statistics (Mean, SD, Median) were reported for all variables. Assumption of normality was checked with the use of Q-Q plots. The main statistical analyses for the first three hypotheses were nonparametric Kruskal-Wallis tests (α=0.05) with additional paired group comparisons through Dunn’s test (α=0.05) for significant overall effects. To control for multiple testing the Bonferroni correction was used. Effect sizes for the overall Kruskal-Wallis tests were expressed in eta-squared (η^2) and effect sizes for paired group comparisons were expressed in ‘r’. Eta-squared effect size was classified as small (0.01), medium (0.06) or large (0.14). Effect size ‘r’ was classified as small (0.10), medium (0.30), or large (0.50). Kruskal-Wallis tests of the first assessments were conducted to check the extent to which the measures derived from the tests used in this study could discriminate the groups. The Kruskal-Wallis tests of the variability scores were conducted to test if there were differences between the groups in terms of how (in)consistently they performed. The additional statistical analysis for the fourth hypothesis was a hierarchical logistic regression analysis (α=0.05). This analysis was conducted to see if the addition of variability scores would provide incremental validity. Regardless of assumptions a non-parametric approach was continued here to be able to compare results accurately and to align with the non-parametric nature of the Kruskal-

Wallis tests. The following eight variables were added at Block 1: RSD-1, GETINDEX-1, GETERR-1, GETRT-1, WAFVMISS-1, WAFVFALSE-1, WAFVRT-1 and WAFVRTSD-1. The following eight variables were added at Block 2: RSD-C, GETINDEX-C, GETERR-C, GETRT-C, WAFVMISS-C, WAFVFALSE-C, WAFVRT-C, and WAFVRTSD-C. In this regression the dependent variable (DV) was 'group', which had a range of 1 (patient group) or 2 (simulation group). The variables that were added at block 1 and block 2 respectively, were the independent variables (IV) that served as predictors in the model. The amount of gained explained variance through the variables included at Block 2, over the amount of explained variance at Block 1 was expressed in Cox & Snell's 'R^2'.

Results

Group differences at T1

All results are presented in table 2 and table 3. Omnibus Kruskal-Wallis test revealed significant group differences (α=0.05), with large effect sizes, for all T1-variables except for the GET mean reaction time (GETRT-1). Bonferroni correction for multiple testing was applied. Additional paired group comparisons (α=0.05) for the significant omnibus results indicated that the control group did not differ significantly from the patient group. The group comparisons also indicated that the simulation group scored significantly lower (α=0.05) than the control group for the RSD-1, on all the other T1 variables the simulation group scored significantly higher than the control group, all with large effect sizes. Lastly, the group comparisons indicated that the simulation group scored significantly lower than the patient group (α=0.05) for the RSD-1 and scored significantly higher than the patient group for the GETINDEX-1, GETERRORS-1, WAFVMISS-1, and the WAFVFALSE-1, all with medium to large effect sizes.

Group differences in consistency

Omnibus Kruskal-Wallis test revealed significant group differences (α=0.05), with medium to large effect sizes for three out of eight consistency variables (GETERR-C, WAFVMISS-C, and WAFVFALSE-C). Bonferroni correction for multiple testing was applied. Additional paired group comparisons (α=0.05) for the significant omnibus results indicated that the control group did not differ significantly from the patient group. Group comparisons also indicated that the simulation group scored significantly higher (α=0.05) than the control group on two of the variables with significant omnibus results, GETERR-C and WAFVFALSE-C, with medium effect sizes. The simulation group also scored significantly higher than the patient group (α=0.05) on the variables WAFVMISS-C and WAFVFALSE-C, with medium to large effect sizes.

Incremental value of repeated assessment and variability scores

The T1-variables were included in the model at Block 1. These are the following eight variables: RSD-1, GETINDEX-1, GETERR-1, GETRT-1, WAFVMISS-1, WAFVFALSE-1, WAFVRT-1 and WAFVRTSD-1. The consistency variables were added at Block 2. These are the following eight variables: RSD-C, GETINDEX-C, GETERR-C, GETRT-C, WAFVMISS-C, WAFVFALSE-C, WAFVRT-C, and WAFVRTSD-C. The hierarchical logistic regression analysis indicated a significant effect at block 2 when adding the consistency variables (R^2=0.740, χ^2=16.489, df= 16, p = .036, α=0.05; Block 1: R^2=0.625, χ^2=44.082, p<.001, α=0.05). This effect shows a 12% gain in predictive value of the model when including the consistency variables at block 2. In this regression the dependent variable (DV) was ‘group’, which had a range of 1 (patient group) or 2 (simulation group). The variables that were added at block 1 and block 2 respectively, were the independent variables (IV) that served as predictors in the model.

Table 2

Descriptive Statistics of PVT-Results per Group

Variable	Controlgroup					Patientgroup					Simulationgroup				
	Mean	Median	SD	Min	Max	Mean	Median	SD	Min	Max	Mean	Median	SD	Min	Max
RSD-1	10.69	10.00	2.24	8.00	15.00	9.78	9.00	2.16	700	14.00	7.70	8.00	2.11	4.00	13.00
GETINDEX-1	-1.31	-1.59	4.57	-11.31	10.38	-0.43	-0.52	1.87	-3.06	3.92	4.16	3.81	4.83	-2.26	17.23
GETERR-1	6.06	4.00	8.31	0.00	28.00	6.39	5.00	5.53	0.00	16.00	16.11	14.00	10.57	1.00	48.00
GETRT-1	4.40	3.64	1.98	1.89	9.88	4.81	3.84	2.87	2.18	14.36	3.51	3.45	1.27	1.21	5.62
WAFVMISS-1	1.13	0.00	1.63	0.00	5.00	0.83	0.00	1.30	0.00	4.00	6.59	5.00	6.07	0.00	23.00
WAFVFALSE-1	1.38	1.00	1.15	0.00	3.00	2.67	2.00	2.87	0.00	10.00	10.22	9.00	7.72	1.00	32.00
WAFVRT-1	398.00	378.75	63.45	334.00	529.00	439.69	440.00	79.26	300.00	598.00	515.13	498.50	142.80	290.00	843.00
WAFVRTSD-1	1.21	1.21	0.05	1.11	1.28	1.23	1.24	0.07	1.10	1.39	1.29	1.31	0.09	1.17	1.48
RSD-C	1.50	1.00	0.97	0.00	3.00	1.28	1.00	1.23	0.00	5.00	1.89	2.00	1.63	0.00	6.00
GETINDEX-C	3.03	2.28	2.70	0.23	10.99	1.71	1.27	1.29	0.10	4.23	2.51	1.87	2.52	0.07	11.55
GETERR-C	3.38	1.00	5.30	0.00	19.00	2.89	2.50	2.91	0.00	11.00	6.14	4.00	5.23	0.00	18.00
GETRT-C	2.19	1.80	1.67	0.34	6.31	1.64	1.47	1.07	0.04	4.25	1.46	1.30	1.24	0.00	4.53
WAFVMISS-C	2.13	1.50	2.42	0.00	8.00	0.94	0.50	1.26	0.00	4.00	3.52	3.00	2.62	1.00	9.00
WAFVFALSE-C	2.44	1.00	2.99	0.00	8.00	2.28	1.50	2.72	0.00	9.00	7.52	5.00	8.39	0.00	41.00
WAFVRT-C	74.28	53.00	84.66	12.00	316.00	58.42	33.50	86.40	0.00	354.00	92.43	57.50	94.19	5.50	356.00
WAFVRTSD-C	0.08	0.06	0.09	0.01	0.26	0.06	0.06	0.05	0.01	0.17	0.10	0.09	0.08	0.00	0.38

Note. Abbreviations: RSD-1=Reliable Digit Span at T1; GETINDEX-1=GET Index score at T1; GETERR-1 =GET Errors at T1; GETRT-1 =GET mean response time at T1 expressed in ….; WAFVMISS-1=WAFV Omission errors at T1; WAFVFALSE-1=WAFV Commission errors at T1; WAFVRT-1 =WAFV mean response time at T1 expressed in milliseconds; WAFVRTSD-1=WAFV standard deviation of mean response time at T1. Pairwise comparisons were only conducted for significant effects of Kruskal-Wallis test; RSD-C =absolute difference score of Reliable digit span scores at T1 and T2; GETINDEX-C =absolute difference score of GET index scores at T1 and T2; GETERR-C =absolute difference score of GET errors at T1 and T2; GETRT-C =absolute difference score of GET mean response times at T1 and T2; WAFVMISS-C =absolute difference score of WAFV omission errors at T1 and T2; WAFVFALSE-C =absolute difference score of WAFV commission errors at T1 and T2; WAFVRT-C= absolute difference score of WAFV mean response times at T1 and T2; WAFVRTSD-C =absolute difference score of WAFV standard deviations of mean response times at T1 and T2.

Table 3

Kruskal-Wallis Results and Paired Group Comparison Results

Variable:	H (2, n=61)	p-value	η^2	Paired Group Comparisons								
				SIMULATORS – CONTROLS			SIMULATORS – PATIENTS			PATIENTS - CONTROLS		
				H(2, n=61)	p-value	r	H(2, n=61)	p-value	r	H(2, n=61)	p-value	r
RSD-1	18.09	<.001*	.276	22.133	<.001*	.610	15.41	.012*	.430	6.726	.794	.191
GETINDEX-1	18.69	<.001*	.288	-21.733	<.001*	.592	-17.58	.003*	.485	-4.149	1.000	.117
GETERR-1	17.82	<.001*	.273	-20.898	<.001*	.570	-17.58	.003*	.484	-3.398	1.000	.096
GETRT-1	2.92	.232	.016									
WAFVMISS-1	22.87	<.001*	.360	-20.414	<.001*	.571	-22.02	<.001*	.624	1.604	1.000	.046
WAFVFALSE-1	29.77	<.001*	.479	-27.530	<.001*	.755	-21.59	<.001*	.600	-5.937	.981	.168
WAFVRT-1	12.03	.002*	.173	-19.329	.002*	.526	-8.93	.295	.246	-10.40	.264	.292
WAFVRTSD-1	10.42	.005*	.145	-17.340	.006*	.473	-10.94	.128	.302	-6.396	.881	.180
RSD-C	2.11	.349	.002									
GETINDEX-C	2.44	.295	.008									
GETERR-C	7.92	.019*	.102	-14.683	.025*	.403	-10.23	.168	.284	-4.451	1.000	.126
GETRT-C	2.54	.281	.009									
WAFVMISS-C	14.94	<.001*	.223	-11.382	.115	.316	-20.19	<.001*	.568	8.813	.423	.253
WAFVFALSE-C	12.90	.002*	.188	-16.324	.010*	.449	-16.24	.007*	.452	-0.083	1.000	.002
WAFVRT-C	3.29	.193	.022									
WAFVRTSD-C	4.24	.120	.039									

Note. Abbreviations: RSD-1 = Reliable Digit Span at T1; GETINDEX-1 = GET Index score at T1; GETERR-1 = GET Errors at T1; GETRT-1 = GET mean response time at T1 expressed in ….; WAFVMISS-1= WAFV Omission errors at T1; WAFVFALSE-1 = WAFV Commission errors at T1; WAFVRT-1 = WAFV mean response time at T1 expressed in milliseconds; WAFVRTSD-1 = WAFV standard deviation of the mean response time at T1. Pairwise comparisons were only conducted for significant effects of the Kruskal-Wallis test; RSD-C = absolute difference score of the Reliable digit span scores at T1 and T2; GETINDEX-C = absolute difference score of the GET index scores at T1 and T2; GETERR-C = absolute difference score of the GET errors at T1 and T2; GETRT-C = absolute difference score of the GET mean response times at T1 and T2; WAFVMISS-C = absolute difference score of the WAFV omission errors at T1 and T2; WAFVFALSE-C = absolute difference score of the WAFV commission errors at T1 and T2; WAFVRT-C = absolute difference score of the WAFV mean response times at T1 and T2; WAFVRTSD-C = absolute difference score of the WAFV standard deviations of the mean response times at T1 and T2. Pairwise comparisons were only conducted for significant effects of the Kruskal-Wallis test.
** p-values that indicate significant effects are highlighted with an asterisk.*

Discussion

This study was designed to evaluate the potential incremental value of using scores of variability as measure for (in)consistency in performance on repeated administration of PVTs to detect feigned ADHD in adulthood in clinical assessments. The first main result of this study is that at a single assessment, all PVTs discriminated well between genuine patients with ADHD and individuals feigning ADHD. The second main result is that the use of consistency scores did not discriminate well between genuine patients with ADHD and individuals feigning ADHD. Only for a few measures did the consistency scores show differences between those two groups. Lastly, this study found that the use of consistency scores in a model already including scores from a single assessment did not add a significant amount of incremental value for discriminating genuine patients with ADHD and individuals feigning ADHD. Taken together, the main implications of these findings are that single assessment PVTs are successful in discriminating genuine patients with ADHD from individuals feigning ADHD and a single assessment using PVTs to be sufficient for detecting feigned ADHD in clinical assessment. Furthermore, consistency scores from repeated assessments using PVTs do not distinguish well between the two groups, therefore not supporting the use of repeated assessment with PVTs in the clinical assessment of ADHD in adults. And lastly, not only do consistency scores not appear useful for detecting feigned ADHD as separate method, they also do not add unique additional value to detect feigned ADHD when used alongside single assessment scores. This suggests that any potential use for repeated assessment and consistency scores in detecting feigned ADHD has yet to be supported by research.

The first main result of the study, regarding the successful differentiation between genuine patients with ADHD and individuals feigning ADHD using a single assessment including PVTs, supports the basis for all hypotheses of this research. This result shows that

the tests that were used to derive the consistency scores from work in the ways that were expected. They were able to successfully distinguish genuine ADHD patients from simulating individuals, and to distinguish genuine healthy individuals from simulating individuals. There were however non-significant comparisons between the genuine ADHD patients and healthy individuals. This was not in line with expectations, as it was expected that healthy individuals would perform significantly different from patients with ADHD. A possible explanation for this unexpected result could be the small sample size of the control group (n=16), also relative to the larger size of the simulation group (n=27). Another possible interpretation of this result could be that, taken the small sample size into account, this research included a control group that scored much higher than common for healthy individuals on these tests. The ways that the other two groups are successfully distinguished from each other make this explanation more plausible, as a fault in these tests would have to affect the other results as well but they were unaffected. It was expected for PVTs in a single assessment to be a successful detection method for feigned ADHD, as has also been shown in previous research (Fuermaier et al., 2022; Rogerset al., 2021). The commission and omission errors of the WAFV presented as the most useful test measures for distinguishing genuine ADHD patients from individuals feigning ADHD, with largest effect sizes.

The only measure that did not show significant group differences at the omnibus Kruskal-Wallis test of the single assessment scores was the GETRT-1. This was not aligned with the expectation that all the measures would be able to distinguish genuine ADHD patients and individuals feigning ADHD. In research of Fuermaier et al. (2020), in which patients with acquired brain injury (ABI), healthy control participants and participants instructed to simulate ABI performed the GET, there were significant differences in mean reaction time between all groups. This could suggest that a parallel between research with other groups, such as an ABI group and an ADHD group cannot be drawn directly. Another

possible explanation could be that participants of the simulation group in this study did not focus on aspects of reaction time as much as they focused on other aspects of feigning like making certain kinds of mistakes, but this hypothesized explanation has not been thoroughly investigated. Research about feigned ADHD that includes feigning strategy often translates the feigning strategy on a performance-based test in terms of 'types' of symptom-feigning, such as in research of Sullivan et al. (2007). In this study the authors consider underperformance on a performance-based validity test, the Word Memory Test (WMT), to be a form of symptom exaggeration. However, this does not shed light on the ways in which individuals that simulate ADHD symptoms through performance do so in terms of their behaviors. The research does not address if they intuitively improvise during their simulating, or if they are intentional in terms of which kind of mistakes they decide to make, or how many in total or whether the timing of their mistakes is actively considered during the testing. The current study also did not include feigning strategy, future studies may want to do so to be able to control for this.

The second main result of the study, that consistency scores derived from repeated assessment using PVTs did not discriminate between genuine patients with ADHD and individuals feigning ADHD better than scores of a single assessment, was not in alignment with the first three hypotheses of this study. It was hypothesized that (1) the control group would show the smallest absolute variability and thus the most consistent performance, (2) the patient group would show larger absolute variability scores than the control group, but smaller absolute variability scores than the simulation group, thus performing less consistent than the control group but more consistent than the simulation group, (3) the simulation group may show the largest absolute variability scores of all the groups and thus show the least consistent performance. The results did not show that the groups performed in these ways. The third main result of the study shows the lacking incremental value of consistency scores in a model

alongside single assessment scores for distinguishing between genuine patients with ADHD and individuals feigning ADHD. This finding was not in alignment with the fourth hypothesis of this study. It was hypothesized that (4) the differences between the groups may be larger for the variability scores than the differences between the groups at the first assessment, thus showing the incremental value of repeated assessment and variability scores. However, consistency variables to a model increased its' predictive value with 12% compared to a model with only single-assessment scores, which is only a very small proportion that is uniquely explained by using the consistency variables. Based on theories such as the medical model, in which symptoms are defined as signals of underlying problems that are stable and remain permanent unless treated (Blaney, 2014), it was expected that the consistency scores would be at least equally successful in distinguishing the two groups, if not better, and that the consistency method would explain a unique part of feigning in ADHD assessments. As these last two main results from this study did not support these expectations, this also contradicts the general public's assumption that individuals who feign symptoms are unsuccessful at this over longer periods of time (Jelicic et al.,2017). The last two main results of this study align with previous research of Boskovic and colleagues (2022) in which it was also found that it was not consistency that aids in detection of feigned symptoms. The authors found that overreporting of symptoms was a better aid in detection of feigned symptoms. A limitation of their study was that five days may not have been a long enough period to measure the intraindividual variability, which the current study has attempted to control by lengthening of the time frame between assessments. The results, however, do not seem to have changed by doing so. This could suggest that larger time frames will not show inconsistency because it does not aid in the detection of feigning. The results of the current research also do not show that using consistency scores on top of single assessment scores cause enough incremental value to justify the time and effort needed for repeated assessments in practice. Taken

together, and placed into the current body of literature, the findings from this study do not seem to be promising research ventures regarding the consistency method in the context of detection methods for feigning in ADHD assessments. The debated nature of whether the consistency assumption holds true or not, and specifically in ADHD symptomology, remains (Peace et al., 2010).

This research needs to be discussed in the context of several limitations. First, this study relied on a convenience sampling method. This may have affected the nature of the sample in such a way that any demographic characteristics of the researcher have extended to the participants, as they were recruited from the social network of the researcher. This may have resulted in a rather homogenous sample, with regards to demographic characteristics. Second, as no incentive was offered to the sample recruited for the simulation and control groups, in the way that a financial reward was offered to the patients that participated in the patient group, this study cannot have adequately imitated real-life benefits of feigning. This limits the ecological validity of the study. Third, feigning scenario and instructions may have not been sufficiently understood by the participants. The only way there was checked for this was through the pre- and post-experimental check questionnaires. This may have been a lack of ways to allow participants to show the extent to which they really understood the concepts and instructions, which may have hindered them to feign according to their most sophisticated and convincing efforts. Another phenomenon that could have influenced results is the power of the study. A power analysis was not performed prior to this study. As mentioned in research of Fuermaier et al. (2020), in simulation studies power issues are usually no great concern because of the relatively large effects that are needed to detect feigning. This can be revealed even with small samples, as Rogers (2008) shows. He states that based on a two-tailed group comparison at α=.05, a power (1-β) of .85 requires 33 participants per group as sample size for a moderate effect (d=.75) and only 13 participants per group for a large effect

(d=1.25). This study employed a control group of 16 participants, a patient group of 18 participants and a simulation group of 27 participants, so at face value power issues do not seem to underly non-significant results. It should be mentioned however, that a rather small control group may have been an underlying cause for the non-significant results of the paired group comparisons between the patient group and control group at the single assessment scores. Another aspect of this research that was affected by a small sample size was the logistic regression analysis. Combined with many predictor variables at both blocks of the analysis, the small sample size may have caused an underpowered analysis and affected the result. Lastly, a repeated assessment consisting of two assessments may not have captured the way in which performance of patients with ADHD truly fluctuates. Over much longer periods of time, with more assessment moments, intraindividual response variations may show much more clearly.

Future studies may want to invest in recruitment of larger and more heterogeneous samples and to conduct a power analysis prior to commencing the research. They may also want to employ an approach of simulation design in which response patterns of patients are measured multiple times per day, over a longer timeframe and thereafter compared to instructed feigners. This may capture the fluctuating nature of how ADHD manifests in performance much more accurately, allowing long-term trends in performance to potentially appear.

References

American Psychological Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*: DSM–5. Washington, DC: American Psychological Association.

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain related disability: Lessons from cognitive malingering research. The Spine Journal: Official Journal of the North American Spine Society, 5(4), 404–417. https://doi.org/10.1016/j. spinee.2004.11.016

Blaney, P. H. (2014). Medical model of mental disorders. In Cautin, R. L. & Lilienfeld, S. O. (Eds.), *The Encyclopedia of Clinical Psychology (pp. 1–6)*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118625392.wbecp382

Boskovic, I., Zwaan, L., Baillie, V., & Merckelbach, H. (2022). Consistency does not aid detection of feigned symptoms, overreporting does: Two explorative studies on symptom stability among truth tellers and feigners. *Applied Neuropsychology: Adult, 29:6*, 1458-1466, DOI: 10.1080/23279095.2021.1888728

Braun, D., Dulit, R., Adler, D., Berlant, J., Dixon, L., Fornari, V., & Richter, D. (2004). Attention-deficit hyperactivity disorder in adults: Clinical information for primary care physicians. *Primary Psychiatry, 11*, 56– 65.

Conners, K., Erhardt, D., & Sparrow, E. (1999). Conners Adult ADHD Rating Scales Technical Manual. Tonawanda, New York: Multi-Health Systems, Inc.

Conners, C. K. (2014). Conners' Continuous Performance Test – 3rd Edition (CPT-3) manual. Toronto, ON: Multi-Health Systems.

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

Dong, H., Koerts, J., Pijnenborg, G.H.M., Scherbaum, N., Müller, B.W., & Fuermaier, A.B.M. (2023). Cognitive Underperformance in a Mixed Neuropsychiatric Sample at Diagnostic Evaluation of Adult ADHD. *J. Clin. Med.* 2023, 12, 6926. https://doi.org/10.3390/jcm12216926

DuPaul, A., Power, T., Anastopoulos, A., & Reid, R. (1998). ADHD rating scale IV: Checklists, norms, & clinical interpretation. New York: Guilford Press.

Faraone, S. V., Biederman, J., Feighner, J. A., & Monuteaux, M. C. (2000). Assessing symptoms of attention deficit hyperactivity disorder in children and adults: Which is more valid? *Journal of Consulting and Clinical Psychology, 68*, 830 – 842. http://dx.doi.org/10.1037/0022006X.68.5.830

Fisher, A., & Watkins, M. (2008). ADHD rating scales–susceptibility to faking in a college student sample. Journal of Postsecondary Education and Disability, 20, 81–92.

Fuermaier, A. B. M., Tucha, L., Koerts, J., Aschenbrenner, S., & Tucha, O. (2017). Vienna Test System (VTS): The Groningen Effort Test (GET) for the assessment of noncredible cognitive performance in the assessment of adult ADHD. Vienna, Austria: Schuhfried.

Fuermaier, A. B. M., Tucha, O., Koerts, J., Grabski, M., Lange, K. W., Weisbrod, M., Aschenbrenner, S., Tucha, L. (2016) The Development of an Embedded Figures Test for the Detection of Feigned Attention Deficit Hyperactivity Disorder in Adulthood. Plos one, 11(10): e0164297. doi:10.1371/journal.pone.0164297

Fuermaier, A.B.M., Tucha, O., Russ, D., Ehrenstein, J.K., Stanke, M., Heindorf, R., Buggenthin, R., Aschenbrenner, S., Koerts, J., & Tucha, L., (2020). Utility of an attention-based performance validity test for the detection of feigned cognitive dysfunction after acquired brain injury, Journal of Clinical and Experimental Neuropsychology, 42:3, 285-297, DOI: 10.1080/13803395.2019.1710468

Fuermaier, A. B. M., Tucha, L., Guo, N., Mette, C., Müller, B. W., Scherbaum, N., & Tucha, O. (2022). It Takes Time: Vigilance and Sustained Attention Assessment in Adults with ADHD. International Journal of Environmental Research and Public Health, 19(9), Article 5216. https://doi.org/10.3390/ijerph19095216

Granhag, P. A., & Stromwall, L. A. (1999). Repeated interrogations stretching the deception detection paradigm. *Expert Evidence, 7*(3), 163–174. https://doi.org/10.1023/A:1008993326434

Greenfield, B., Ochs, E., & Hechtman, L. (2002). Diagnosing attention deficit-hyperactivity disorder in adults: Challenges and controversies. *Drug Benefit Trends, 14*, 27–33.

Greiffenstein, M. F., Baker, W.J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological assessment*, *6*, 218.

Greve, K. W., Springer, S., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Ciota, M. A. (2007). Malingering in toxic exposure: Classification accuracy of reliable digit span and WAIS-III digit span scaled scores. Assessment, 14(1), 12–21.

Greve, K. W., Bianchini, K. J., & Brewer, S. T. (2013). The assessment of performance and self-report validity in persons claiming pain related disability. *The Clinical Neuropsychologist, 27*(1), 108–137. https://doi.org/10.1080/13854046.2012.739646

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

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

Harrison, A. G., Rosenblum, Y., & Currie, S. (2010). Examining unusual digit span

performance in a population of postsecondary students assessed for academic difficulties. *Assessment*, 17, 283–293. http://dx.doi.org/10.1177/1073191109348590

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

Jachimowicz, G., & Geiselman, R. (2004). Comparison of ease of falsification of attention-deficit/hyperactivity disorder diagnosis using standard behavior rating scales. *Cognitive Science Online, 2*, 6 –20.

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

Jelicic, M., Merckelbach, H., & Boskovic, I. (2017). Seven myths about feigning. In H. Otgaar & M. L. Howe (Eds.), Finding the truth in the courtroom: Handling deception, lies, and memories. Oxford University Press.

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.proxy-ub.rug.nl/10.1037/pas0000247

McDermott, B. E., & Feldman, M. D. (2007). Malingering in the medical setting. *Psychiatric Clinics of North America, 30*(4), 645–662. https://doi.org/10.1016/j.psc.2007.07.007

McGough, J. J., & Barkley, R. A. (2004). Diagnostic controversies in adult attention deficit hyperactivity disorder. *The American Journal of Psychiatry*, *161*(11), 1948–1956.

https://doi-org.proxy-ub.rug.nl/10.1176/appi.ajp.161.11.1948

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*(8), 1094–1102. https://doi.org/10.1076/jcen.24.8.1094.8379

Murphy, K. R., & Adler, L. A. (2004). Assessing attention-deficit/hyperactivity disorder in adults: Focus on rating scales. *The Journal of Clinical Psychiatry, 65(Suppl. 3)*, 12–17.

Musso, M. W., & Gouvier, W. (2014). "Why is this so hard?" A review of detection of malingered ADHD in college students. *Journal of Attention Disorders, 18*, 186-201.

Peace, K. A., Porter, S., & Cook, B. L. (2010). Investigating differences in truthful and fabricated symptoms of traumatic stress over time. *Psychological Injury and Law, 3*(2), 118–129. https://doi.org/10.1007/s12207-010-9078-8

Petermann, F. (2012). Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV). German version. Frankfurt: Pearson Assessment.

Primich, C., & Iennaco, J. (2012). Diagnosing adult attention-deficit hyperactivity disorder: The importance of establishing daily life contexts for symptoms and impairments. *Journal of Psychiatric and Mental Health Nursing*, *19*(4), 362–373. https://doi-org.proxy-ub.rug.nl/10.1111/j.1365-2850.2011.01845.x

Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology, 18*, 379 –395. http://dx.doi.org/10.1093/arclin/18.4.379

Rios, J., & Morey, L. (2013). Detecting feigned ADHD in later adolescence: An examination of three PAI negative distortion indicators. *Journal of Personality Assessment, 95*, 594 –599.

Robinson, A., Calamia, M., Penner, N., Assaf, N., Razvi, P., Roth, R. M., & Erdodi, L. A. (2023). Two times the charm: Repeat administration of the CPT-II improves its

classification accuracy as a performance validity index. *Journal of Psychopathology and Behavioral Assessment*, *45*(3), 591–611. https://doi-org.proxy-ub.rug.nl/10.1007/s10862-023-10055-7

Rogers, R. (2008). Clinical assessment of malingering and deception. New York, NY: Guilford Press.

Rogers, R., Velsor, S. F., Donnelly, J. W., 2nd., & Dean, B. (2021). Embedded WAIS-IV detection strategies and feigned cognitive impairment: An investigation of malingered ADHD. Assessment, 28(1), 44–56. https://doi.org/10.1177/1073191120927788

Schuhfried, G. Vienna Test System (VTS) 8 (Version 8.2.00) [Computer Software]; SCHUHFRIED: Mödling, Austria, 2013.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *Clinical Neuropsychologist, 13*, 545–561. http://dx.doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545

Suhr, J. A., & Berry, D. T. R. (2017). The importance of assessing for validity of symptom report and performance in attention deficit/hyperactivity disorder (ADHD): Introduction to the special section on noncredible presentation in ADHD. *Psychological Assessment, 29*, 1427–1428. http://dx.doi.org/10.1037/pas0000535

Suhr, J. A., Buelow, M., & Riddle, T. (2011). Development of an Infrequency Index for the CAARS. *Journal of Psychoeducational Assessment, 29*, 160 –170. http://dx.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, 189 –207. http://dx.doi.org/10.1080/09084280701509083

Tombaugh, T. N. (1996). Test of memory malingering. Toronto, Canada: Multi-Health

Systems.

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. https://doi-org.proxy-ub.rug.nl/10.1007/s00702-014-1274-3

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. http://dx.doi.org/10.1007/s12402-009-0007-1

van der Heide, D., Boskovic, I., van Harten, P., & Merckelbach, H. (2020). Overlooking feigning behavior may result in potential harmful treatment interventions: two case reports of undetected malingering. *Journal of Forensic Sciences, 65*(4), 1371–1375. https://doi.org/10.1111/1556-4029.14320

Wallace, E. R., Garcia-Willingham, N. E., Walls, B. D., Bosch, C. M., Balthrop, K. C., & Berry, D. T. R. (2019). A meta-analysis of malingering detection measures for attention-deficit/hyperactivity disorder. *Psychological Assessment*, *31*(2), 265–270. https://doi-org.proxy-ub.rug.nl/10.1037/pas0000659.supp

Wei, C., & Suhr, J. A. (2015). Examination of the role of expectancies on task performance in college students concerned about ADHD. *Applied Neuropsychology: Adult, 22*, 204 – 208. http://dx.doi.org/10.1080/23279095.2014.902836

Williamson, K. D., Combs, H. L., Berry, D. T. R., Harp, J. P., Mason, L. H., & Edmundson, M. (2014). Discriminating among ADHD alone, ADHD with a comorbid psychological disorder, and feigned ADHD in a college sample. *Clinical Neuropsychologist, 28*, 1182–1196. http://dx.doi.org/10.1080/13854046.2014.956674