Applied Neuropsychology: Adult

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hapn21

Discriminating malingered attention Deficit Hyperactivity Disorder from genuine symptom reporting using novel Personality Assessment Inventory validity measures

Allyson G. Harrison, Kathleen A. Harrison & Irene T. Armstrong

To cite this article: Allyson G. Harrison, Kathleen A. Harrison & Irene T. Armstrong (2022) Discriminating malingered attention Deficit Hyperactivity Disorder from genuine symptom reporting using novel Personality Assessment Inventory validity measures, Applied Neuropsychology: Adult, 29:1, 10-22, DOI: 10.1080/23279095.2019.1702043

To link to this article: https://doi.org/10.1080/23279095.2019.1702043

APPLIED NEUROPSYCHOLOGY: ADULT
2022, VOL. 29, NO. 1, 10–22
https://doi.org/10.1080/23279095.2019.1702043

Routledge
Taylor & Francis Group

Discriminating malingered attention Deficit Hyperactivity Disorder from genuine symptom reporting using novel Personality Assessment Inventory validity measures

Allyson G. Harrison, Kathleen A. Harrison, and Irene T. Armstrong

Department of Psychology, Queen's University (Regional Assessment and Resource Centre), Kingston, Canada

ABSTRACT

It is now widely understood that ADHD can be feigned easily and convincingly. Despite this, almost no methods exist to assist clinicians in identifying when such behavior occurs. Recently, new validity indicators specific to feigned ADHD were reported for the Personality Assessment Inventory (PAI). Derived from a logistic regression, these algorithms are said to have excellent specificity and good sensitivity in identifying feigned ADHD. However, these authors compared those with genuine ADHD only to nonclinical undergraduate students (asked to respond honestly or asked to simulate ADHD); no criterion group of definite malingerers was included. We therefore investigated these new validity indicators with 331 postsecondary students who underwent assessment for possible ADHD and compared scores of those who were eventually diagnosed with ADHD ($n = 111$) to those who were not [Clinical controls (66), Definite malingerers (36); No diagnosis (117)]. The two proposed PAI algorithms were found to have poor positive predictive value (.19 and .17). Self-report validity measures from the Connors' Adult Attention Rating Scale, and the Negative Impression Management scale on the PAI returned more positive results. Overall, more research is needed to better identify noncredible ADHD presentation, as the PAI-based methods proposed by Aita et al. appear inadequate as symptom validity measures.

KEYWORDS

ADHD; Symptom Validity Testing; non-credible performance; adult

Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a common neurodevelopmental disorder that affects up to 9.4% of school-aged children (American Psychiatric Association; APA, 2013; Danielson et al., 2018) and up to 4.4% of adults in North America (Kessler et al., 2006). The inattention, hyperactivity and/or impulsivity symptoms associated with this disorder significantly impair individuals in two or more major life areas (APA, 2013). While it is relatively easy to diagnose this condition in children due to ease of observation and access to multiple informant ratings from those who interact regularly with the child, it is much more difficult to diagnose in adults (Greenfield, Ochs, & Hechtman, 2002; McGough & Barkley, 2004). This is due to a number of factors, including problems obtaining accurate retrospective information about childhood symptoms (Mannuzza, Klein, Klein, Bessler, & Shrout, 2002), high prevalence rate of inattentive symptoms in many other psychological conditions (Alexander & Harrison, 2013; Harrison, 2004; Harrison, Alexander, & Armstrong, 2013), and ease of feigning symptoms of ADHD by adults motivated to obtain the diagnosis for reasons of secondary gain (Harrison, Edwards, & Parker, 2007; Jachimowicz & Geiselman, 2004; Marshall, Hoelzle, Heyerdahl, & Nelson, 2016; Musso & Gouvier, 2014; Sullivan, May, & Galbally, 2007).

Issues in diagnosis of ADHD in post-secondary students

In adults, a clinical diagnosis of ADHD is often made based mainly on interview data and symptom report checklists (Jachimowicz & Geiselman, 2004) and such methods are extremely susceptible to feigned impairment. Indeed, multiple studies (e.g., Booksh, Pella, Singh, & Gouvier, 2010; Harrison et al., 2007; Jachimowicz & Geiselman, 2004; Quinn, 2003; Sollman, Ranseen, & Berry, 2010) have shown how easily both naïve and coached college-aged students can feign believable symptoms on all major ADHD self-report inventories, including exaggeration of historical as well as current symptoms. In postsecondary clinical samples, rates of suspected feigning of ADHD range from 15% to 47.6% (Harrison & Edwards, 2010; Sullivan et al., 2007). These studies suggest that the base rate of feigned ADHD, at least in college-level populations, is higher than the base rate of the condition itself.

Reports in the popular press, too, suggest that enterprising parents and their offspring are gaming the system to ensure that their children are diagnosed with ADHD in an attempt to improve performance on high stakes tests such as the Scholastic Achievement Test (SAT) or the American College Test (ACT). Indeed, over the past decade multiple reports (e.g., Abrams, 2005; Belkin, 2018; Lewak, 2018; Mitchell, 2012; Tapper, Morris, & Setrakian, 2006; Taylor,

CONTACT Allyson G. Harrison harrisna@queensu.ca Department of Psychology, Queen's University (Regional Assessment and Resource Centre), 68 University Ave, Kingston, Canada.

2019) have highlight concerns about the increasingly large number of students from affluent American neighborhoods whose children have been diagnosed as disabled and granted extra time for taking college entrance tests. These reports suggest that students deliberately fail diagnostic tests in order to obtain a disability label and qualify for test accommodations such as extra time. ADHD symptoms may also be fabricated by students as a way to obtain stimulant medication, and multiple studies of college-aged students demonstrate the high rate of stimulant abuse that is occurring at present (e.g., Advokat, Guidry, & Martino, 2008; Hanson et al., 2013; McCabe, Teter, & Boyd, 2006). As such, there is a need to develop objective and reliable methods to detect such symptom exaggeration when it occurs.

Improving identification of feigned ADHD

Unfortunately, very little research to date has identified specific and sensitive methods to identify feigning or exaggeration of ADHD. Harrison, Lee and Suhr (in press) offer a comprehensive review of the extant literature regarding various proposed methods to support diagnosis of ADHD malingering. However, many of these methods have been validated using only simulated malingerers and have not been proven effective in real world assessments. Further, it is unclear whether noncredible ADHD is best identified by use of performance validity measures (PVT) such as those employed in neuropsychological evaluations (Harrison et al., in press) or by symptom validity tests (SVT) which measure a tendency to exaggerate on self-reported measures. Given, however, that ADHD is typically diagnosed by means of self-report (Jachimowicz & Geiselman, 2004), one promising line of research involves SVT-based methods to identify noncredible symptom reporting.

Suhr, Buelow, & Riddle (2011) and Harrison & Armstrong (2016) have both developed methods to identify symptom overreporting on the Conners' Adult Attention Deficit/Hyperactivity Rating Scale (CAARS; Conners, Erhardt, & Sparrow, 1998), a commonly used self-report measure of adult symptoms of ADHD. While both of these research groups developed symptom validity indices with high specificity (CII = .87 and EI = .97), their reported sensitivity to feigned ADHD is lower (CII = .30 and EI = .24).

The Personality Assessment Inventory (PAI; Morey, 1991) is a self-report measure containing various validity indicators that include a number of symptom validity scales that could potentially identify symptom overreporting in those feigning ADHD, including the Negative Impression Management (NIM) and Positive Impression Management (PIM) scales, the Malingering Index (MAL), and the Rogers Discriminant Function (RDF). Until recently, most studies have found that symptom validity scales on commonly employed personality assessment inventories are insensitive to feigned ADHD (see Kirk, Hutaff-Lee, Connery, Baker, & Kirkwood, 2014; Sullivan et al., 2007). The adolescent version of the PAI, however, was found to be effective in detecting 17- or 18-year-old ADHD simulators when compared to a clinical standardization sample described in the PAI-A manual (Rios & Morey, 2013), but these findings have yet to be replicated in an actual clinical sample. Using a college sample and archival clinical data, Musso, Hill, Barker, Pella, and Gouvier (2016) investigated alternative cutoff scores for the adult version of the PAI validity indices (i.e., NIM >= 77, Mal >= 3, RDF >= 1). While they found high specificity rates for embedded malingering indices, the sensitivity of these was low for ADHD simulators. As such, research to date suggests that self-report instruments like the PAI may be informative as indicators of feigned ADHD, but alternative cutoff scores for utility in psychological assessments are likely necessary.

Recently, Aita, Sofko, Hill, Musso, & Boettcher (2018) proposed two novel validity measures for the PAI derived from a logistic regression; both scale-based and item-based algorithms were said to have excellent specificity and good to excellent sensitivity in identifying feigned ADHD. The difficulty, however, was that these authors compared those with genuine ADHD to healthy controls or those asked to simulate ADHD (the latter two groups being recruited from undergraduate classes); no criterion group of clinical clients identified as probable or definite malingerers was included. While use of simulators is necessary when a genuine criterion group is either not available or difficult to obtain, it is unclear whether the test scores returned by such simulators accurately capture the behavior of genuine malingerers (Ruiz, Drake, Glass, Marcotte, & van Gorp, 2002; Suhr & Gunstad, 2007). In fact, there is some suggestion that data obtained from undergraduates who participate as simulators may be flawed as these students often invest less than full effort when asked to simulate clinical conditions (An, Zakzanis, & Joordens, 2012). Furthermore, Aita et al. attempted to ensure good effort in their clinical sample retrospectively by means of removing those individuals diagnosed with ADHD in their clinics who failed two or more embedded validity measures from either the Wechsler family of tests (specifically, Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994) which has weak sensitivity to feigned cognitive impairment, see Kirkwood, Hargrave, & Kirk, 2011; Larrabee, 2008; Miele, Gunner, Lynch, & McCaffrey, 2012) and/or at least one of the four validity scale on the PAI (which have unknown ability to identify exaggerated ADHD).

The present study therefore sought to validate both the alternate PAI cutoffs suggested by Musso et al. (2016) for identification of feigned ADHD and the PAI validity algorithms suggested by Aita et al. (2018) by using data from a clinical population. We used archival data to investigate the efficacy of both the Musso et al. and the Aita et al. measures to detect feigning of ADHD in an entire set of clinical data, including those strongly suspected of symptom exaggeration, those who were given an ADHD diagnosis, and treatment-seeking students who thought they might have ADHD but did not receive such a diagnosis. Using this archival dataset, the present study sought to investigate to what extent these proposed methods agree with clinical classifications, and document their specificity, sensitivity, positive and negative

predictive properties when identifying clinical cases where feigned ADHD is strongly suspected or known. This study also set out to compare the efficacy of these novel PAI-based measures with both the existing validity scales of the PAI and also the CAARS-based SVT measures previously described in the literature.

Methods

Participants

Participants in this research-approved study were postsecondary students who completed psycho-educational assessments at a Canadian assessment center between 2010 and 2018. All students were referred to this regional assessment center for a comprehensive assessment to determine whether they had a disability that would require academic accommodations and/or medication to treat their symptoms.

In the present study, we identified all clients in an archival database who sought an assessment due to ADHD as a possible cause of their reported problems, were administered the PAI as part of their assessment, and had agreed to allow their data to be used in research. Exclusion criteria for this investigation included clients who did not consent to have their data used in archival research, did not complete the PAI, were not assessed for possible ADHD, and/or were not administered a PVT during testing.

In total, 331 students (44.4% male; mean age = 22.8 years, $SD = 5.9$ years, range 17–57) were identified who met these criteria. Information pertaining to ethnicity was not collected directly, however, the majority were Caucasian. Diagnosis of ADHD ($n = 111$) was made using all five criteria listed in DSM-IV (APA, 2000) or DSM-5 (APA, 2013), depending on date assessed. Sixty six students were diagnosed with a primary mental health condition and they served as clinical controls. There were also 117 students who underwent assessment for subjective cognitive/academic impairments but were given no formal disability diagnosis, and 36 who were identified as definite malingerers (see below for method of identification). All assessments were conducted by a licensed clinical psychologist or occasionally by a graduate student trainee under the direct supervision of a licensed psychologist.

Materials

Since these assessments were all undertaken to determine not only diagnosis but also possible academic accommodations, the ADHD assessment protocol employed a variety of measures, including: measures of cognitive and academic ability, estimates of self-and observer-reported ADHD symptomatology, a continuous performance test, at least one stand-alone PVT, and a self-report measure of personality. The measures of interest in the present study are described below.

Clinical measures of interest

Personality assessment inventory

The PAI is a widely used, 344-item self-report inventory that measures a number of personality traits. Test-takers read and rate each statement on a four-point Likert scale (1-"Not true at all, False", 2-"Slightly true", 3-"Mainly true", and 4-"Very true"). The test provides the clinician with data on four validity scales (including the NIM and the PIM), 11 clinical scales, 5 treatment scales, and 2 interpersonal scales. The NIM detects exaggerated unfavorable presentation based on bizarre and unlikely symptoms (Morey, 1991). Other PAI metrics of interest include the Roger's Discriminant Function (RDF; Rogers, Sewell, Morey, & Ustad, 1996) and the Malingering Index (MAL; Morey, 1993, 1996). The RDF is a discriminant function designed to detect response patterns inconsistent with clinical populations. It is derived from a combination of discriminant function weighted scores from various PAI scales. Any score greater than zero is said by Rogers et al. (1996) to indicate malingering. The MAL is designed to detect over and under-endorsed items inconsistent with clinical populations; it is derived from eight configural features of various PAI scales with a score of ≥ 5 indicating likely malingering (Morey, 1996). Notably, while both scales were designed to detect overreporting of psychiatric symptoms (in particular, severe psychopathology in the case of MAL), they were not specifically designed to detect overreporting of ADHD symptoms. Musso et al. (2016), however, report that NIM, RDF and MAL showed promise in identifying college students feigning ADHD when using the following cutoffs (NIM >= 77, Mal >= 3, RDF >=1).

In accordance with the recommendations by Morey (1996), subjects with Inconsistency (ICN) scores greater than $T = 72$ and/or those with Infrequency (INF) scores greater than $T = 74$ were removed as their scores were deemed to be invalid due to random or careless responding. Six subjects in our data set has ICN scores over this level, and six had INF scores above this score.

PAI logistic regression equations

As noted above, Aita et al. (2018) developed two logistic regression models based on PAI responses from subjects with genuine ADHD, mood and anxiety disorders, and subjects instructed to deliberately feign ADHD.

To detect which subjects were instructed to deliberately feign ADHD, Aita et al. (2018) developed a logistic regression equation using four PAI subscales: Positive Impression Management (PIM), Schizophrenia – Thought Disorder (SCZ-T), Antisocial Features – Stimulus Seeking (ANT-S), and Depression – Cognitive (DEP-C). Notably, they did not include the NIM in this regression as it had already been used initially to remove those students with a previous ADHD diagnosis who were retrospectively classified as feigning symptoms. Their resulting regression equation, the scale-level Feigned Adult ADHD index (Scale-FAA), correctly identified 46.4% of simulators and 92.3% of subjects

who were diagnosed with ADHD when using a cutoff point of 0.7 or higher. Next, Aita et al. developed a similar regression equation using 24 items from the PAI. This item-level FAA (Item-FAA) regression equation correctly identified 85.2% of simulators and 97% of subjects diagnosed with ADHD when using a cutoff of 0.6 or greater. Both FAA indices reportedly showed substantial improvements in detecting feigned ADHD over existing PAI symptom validity scales. As such, data concerning both scale and specific item responses from existing PAIs were included in the present analysis.

Initially, item-level data from the PAI had not been included in our database. This required manual retrieval of the case files identified, and transfer of the item-level PAI items into the existing database. The final page of the PAI interpretive printout contains the specific question numbers and answers given by a client for each PAI item. For each identified client file, this page was scanned using an automated optical character recognition (OCR) program to allow direct input of participant answers for analysis to eliminate human errors in data input. In addition to internal controls for this OCR program, 18 case files were randomly chosen to ensure accuracy of the program. Here, the data coded by the OCR program was compared with the actual PAI print out in the case file. In each case, accuracy was perfect.

In order to validate that the FAA-item algorithm was inputted and functioning correctly, and to confirm the cut score used, we contacted the first author of the Aita et al. paper. We received validation that the cut score for this formula was .6 or greater, and were also provided with an excel spreadsheet that would calculate the Item-FAA for a client (S. Aita, personal communication, April 5, 2019). We also were instructed that the PAI items in the Aita et al. study were coded as follows: False = 0, Slightly True = 1, Mostly True = 2, and Very True = 3. However, an anonymous reviewer noted that it was critically important to clarify whether PAI item responses (e.g., False to Very True) or item scores (e.g., 0 to 3) were used in originally deriving the item-level regression function. This prompted us to review whether the items used in the equation should be based on response (e.g., False = 0 etc.) or by score (where a False answer may sometimes receive the highest score on a reverse-scored item). Given that seven of the items in the Item-FAA calculation are reverse-scored on the PAI (items 18, 94, 112, 139, 201, 213 and 299), the resulting Item-FAA score changes substantially if the items should be reverse scored. As the Aita et al. paper gave no indication regarding whether items were reverse scored, and the spreadsheet provided by Dr. Aita did not instruct one to reverse-score, we turned to the thesis from which the Aita et al. data had been derived, namely Musso (2013). On page 81 of her thesis, the author clearly states that "it should be noted that Items 235 and 301 are reverse scored per PAI scoring procedure." As such, we agreed with the anonymous reviewer that the seven reverse-scored items in the Item-FAA equation should indeed receive a reverse score (i.e., False = 3, Slightly True = 2, Mostly True = 1, and Very True = 0).

SVTs derived from the CAARS

The CAARS itself is a 66-item self-report scale that allows for the calculation of eight different indices, with some items contributing to more than one scale. CAARS items are rated on a 4-point scale (0 = *not at all*, 1 = *just a little*, 2 = *pretty much*, 3 = *very much*). All students completed the CAARS (self and observer forms), which provide the following scores: a) four factor-derived subscales: b) three scales that correspond to the DSM-IV symptoms of Hyperactivity/Impulsivity, Inattention, and Total DSM symptoms; and c) an overall ADHD Index that is said to measure the "overall level of ADHD symptoms" (Conners et al., 1998, p. 23). The test manual states that the ADHD Index "is the best screen for identifying those '*at risk*' for ADHD" (Conners et al., 1998, p. 23). The manual does not stipulate a specific cutoff score that may be taken to indicate ADHD, but recommends that any score over a *T* value of 65 might be considered to indicate an area of clinically significant problems, and suggests that *T*-scores over 70 or 75 be used as a cutoff for inferring clinically significant problems. In addition, while the manual suggests that individuals obtaining *T*-scores on the ADHD Index of over 70 are likely to meet the diagnostic criteria for ADHD, it also cautions that *T*-scores above 80 on any of the subscales should be considered as possible indicators of symptom exaggeration. Unfortunately, no symptom validity scales currently exist within the CAARS; it includes only an Inconsistency Index designed to identify inconsistent responses to items measuring similar content rather than overreporting of symptoms. Scores on this index of 8 or greater are said to invalidate the obtained CAARS scores due to highly inconsistent responding by the subject.

While not utilized in the original assessments, the current study retrospectively calculated two newly described methods of evaluating self-report credibility on the CAARS so as to compare the Aita et al. (2018) algorithm hit rates to the existing SVTs described by Suhr et al. (2011) and Harrison and Armstrong (2016).

The CAARS Infrequency Index (CII; Suhr et al., 2011) was retrospectively calculated from participant data. The CII is composed of 12 items rarely endorsed by typically developing adults as well as those diagnosed with ADHD. Suhr and colleagues (2011) identified a cut score of >21 as producing few false positive identifications for those with ADHD. The index was found to have modest sensitivity (approximately 30%) and high specificity (approximately 95%). Cook, Bolinger, & Suhr (2016) found that the CII had 52% sensitivity to feigning and 97% specificity for ADHD based on extreme elevations of the three CAARS clinical scales derived from DSM-IV ADHD criteria. Because item-level responses were not recorded for all clients, data from only $n = 230$ clients could be used in the present investigation.

The E-CAARS includes 18 additional symptom validity items embedded within the regular CAARS items, and has been described in detail by Harrison & Armstrong (2016). While allowing for all of the regular CAARS indices to be calculated, the embedded items included in this

experimental version also allow clinicians to identify over-endorsement of ADHD-related symptoms. The sum of these symptom validity items produces a Dissimulation score, while a formula that combines the DI with extreme scores from existing CAARS indices produces the Exaggeration Index (EI). According to Harrison & Armstrong (2016), this Exaggeration Index (EI) is said to have acceptable classification accuracy when discriminating between those feigning ADHD and other clinical groups (including those with ADHD) who were reporting symptoms accurately, with sensitivity to feigning ranging from .24 to .69 and specificity ranging from .74 to .97, depending on cut score used. Since this calculation requires administration of a modified version of the CAARS, not all clinicians employed this version of the test; only $n = 198$ were available in the database.

Performance validity measures

A variety of PVTs were employed in the archival assessments, depending on clinician preference. All assessments included at least two stand-alone PVTs, as this is said to produce excellent specificity when both tests are failed (Odland, Lammy, Martin, Grote, & Mittenberg, 2015). Measured employed in the assessments included at least two of the following:

The Word Memory Test (WMT; Green, 2003) evaluates test-taking compliance. Subjects are given a list of 20 word pairs on a computer screen and must identify all words pairs immediately (Immediate Recall; IR) and after a delay of 30 min (Delayed Recall: DR). The consistency (CNS) of responses between IR and DR is also measured. Following the delayed recognition task, individuals are also asked to identify words from the original list in a multiple choice format (MC); in a paired associates format (PA); and in a free recall format (FR). In accordance with the test manual, scores less than 82.5% on any of the first three subtests were taken to indicate noncredible performance.

The Medical Symptom Validity Test (MSVT; Green, 2004) is a similar, shorter version of the WMT. Like the WMT, the MSVT measures Immediate Recognition (IR) and Delayed Recognition (DR) of a word list, as well as the consistency (CNS) of answers between the two subtests. According to the test manual, scores less than 85 on any of the first three subtests indicate noncredible performance. For both the WMT and MSVT, individuals who demonstrated evidence of a Severe Impairment Profile (SIP; Green, 2008), which might indicate test failure due to a genuine neurological condition rather than poor effort, were removed from further analyses if they passed an additional validity test.

The Test of Memory Malingering (TOMM; Tombaugh, 1996) is a frequently used visual recognition PVT consisting of two learning trials, an immediate retention trial, and a delayed retention trial. Learning trials contain 50 line-drawn pictures, which are paired with a second distractor line drawing during the retention trials. A cutoff of ≤ 45 for both Trial 2 and the retention trial was taken as evidence of noncredible performance.

The Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 1997) requires individuals to learn and recognize a series of "easy" and "hard" number sequences presented at different time intervals on a computer screen. Following the recommendations of Frazier, Frazier, Busch, Kerwood, & Demaree (2008), subjects were deemed to be performing noncredibly if their score on the "hard" items fell below 19.

The Test of Variables of Attention (TOVA; Greenberg, Kindschi, Dupuy, & Hughes, 2007) is a computer-based Continuous Performance Test (CPT). The current version includes a Symptom Exaggeration Index (SEI) developed as a result of research by Leark, Greenberg, Kindschi, Dupuy, & Hughes (2008). A SEI score of 3 or more on the TOVA is said to provide firm evidence of noncredible performance.

Notably, due to an abundance of caution, those who failed only one PVT were not assigned to any group as we could not be sure of the validity of their scores. This includes those individuals with a SIP on the WMT or MSVT who passed additional validity measures.

Procedure

As part of the informed consent process, clients referred for assessment were informed at the first appointment that their data would be entered into a database and used, without any identifying information, in subsequent archival research projects. Clients were told that they did not have to agree to allow their de-identified data to be used in any research study, and that they would still be provided with a complete assessment regardless of their decision. All clients who agreed to have their anonymous data used for this research were included in this ethics-approved study.

Clients underwent a full neuropsychological assessment, including mental health screening surveys and tests of performance validity. Additionally, subjects were asked to provide report cards from childhood, have their parents/caregiver complete a retrospective rating of DSM-based ADHD symptoms (number and frequency) observed in the client prior to age 12, complete both self- and observer-versions of the CAARS, and provide evidence to document substantial impairment in more than one major life activity prior to age 12. This allowed a determination as to whether they met all of the five criteria for diagnosis as outlined in DSM-IV (APA, 2000) or DSM-5 (APA, 2013), depending on year tested.

Classification of malingered neuropsychological dysfunction

Definite malingered neuropsychological dysfunction (MND) was identified using the revised Slick, Sherman, & Iverson (1999) criteria outlined by Slick & Sherman (2012). Identification as definite MND requires both the presence of substantial external incentives and one or more strong indicators of exaggeration/fabrication as well as evidence that these behaviors were not due to psychiatric, neurological or developmental factors. Given the secondary gain potential

from being diagnosed with ADHD (Harrison et al., 2007), we felt that all participants would meet this first criterion. The second criterion requires either below chance performance on one or more forced choice measures, high posterior probability (>= .95) that performance is substantially below actual levels on one or more well-validated psychometric indices, or self-reported symptoms that are unambiguously incompatible with or directly contradicted by directly observed behavior and/or test performance. Using the posterior probabilities reported by Odland et al. (2015), we felt that those failing two or more stand-alone validity tests (out of a total of two or three given, not including validity scales on the PAI or the CAARS) met this second criterion. In addition, the data set included one individual who, after the fact, confessed to deliberate feigning for secondary gain.

Investigation of performance characteristics

Clinically relevant performance characteristics were evaluated using sensitivity, specificity, and positive and negative predictive power. These statistics assist clinicians in determining the likelihood that an individual whose score falls above or below a given cut-score actually belongs to the classification group in question. Sensitivity is the proportion of true positive diagnoses or probability of detection, while specificity is the proportion of true negatives (i.e., the percentage of honest subjects who are correctly identified as not being in the MND group). In order to ensure a low false-positive rate, the accepted convention for most validity research is that specificity should be set at .90 or higher (Boone, 2007; Larrabee, 2012, 2014; Vickery, Berry, Inman, Harris, & Orey, 2001); even though this may reduce sensitivity to malingering, it is felt that this is a reasonable tradeoff to ensure that genuine clients are not falsely accused of malingering. Consequently, a validity test with high specificity is unlikely to be failed by someone with a genuine neurologic dysfunction (Larrabee, 2012).

Positive predictive value (PPV) is the probability that subjects with a score above a certain cut-score are truly exaggerating (MND) while negative predictive value (NPV) is the probability that subjects with a score below a given cutoff truly don't have the condition. Unlike sensitivity and specificity, these values are influenced by the actual base-rate of the disorder in question. As such, and as recommended by Lange & Lippa (2017) we calculated these scores not only using the base rate in the current research sample but also by anticipating estimated base rates found in actual clinical samples (approximately 20% in the study by Marshall et al., 2010, and approximately 30% as found by Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008).

Cutoff scores for all investigated validity methods in this manuscript were examined with receiver operating characteristic (ROC) area under the curve (AUC) analyses. AUCs of 0.7–0.79 are considered acceptable, 0.8–0.89 excellent, and 0.9 or above outstanding (Sussman, Peterson, Connery, Baker, & Kirkwood, 2017).

Table 1. Mean age, Full Scale IQ (FSIQ) and percentage of men as a function of group membership.

		Age		Sex	FSIQ	
Diagnosis	*N*	Mean	*SD*	% Male	Mean	*SD*
ADHD	111	21.9	5.3	63.1	104.2	12.0
Clinical control	66	23.8	5.5	31.8	101.8	12.6
MND	36	23.2	6.6	34.3	96.0	13.7
No diagnosis	117	23.0	6.4	37.6	101.7	15.4

Note. MND: malingered neuropsychological dysfunction.

Results

Population demographics by diagnostic classification

The results of this classification are seen in Table 1, which shows the number of participants per group classification, their mean ages, percentage of men per group, and average Full Scale Intelligence Quotient (FSIQ). There is no difference in mean age among the groups,

$F_{(3,\ 325)} = 1.455$, $p = .23$, mean = 22.8; $SD = 5.9$. The ADHD group was mainly men (63.1%) while the remaining groups were mainly women (clinical control = 68.7%; MND = 66.7%; No Diagnosis = 62.4%), Χ$^2_{(3)} = 24.26$, $p < .001$). Of interest, mean FSIQ was different among the groups, $F_{(3,\ 318)} = 3.017$, $p = 0.03$; post hoc tests showed that mean FSIQ was lowest for the MND group and significantly different from the highest mean FSIQ found for the ADHD group. This finding was not unexpected, given that most cognitive ability scores are artificially deflated in those investing limited effort (Green, Rohling, Lees-Haley, & Allen, 2001).

We investigated the existing PAI validity indices and cut-scores suggested by Musso et al. (2016) and found no difference among groups for inconsistency, infrequency, or RDF (see Tables 2 and 3). Significant differences were found for NIM, and post hoc tests identified that the MND group returned significantly higher scores than all other groups. While significant differences were found among groups on the MAL, (and on the DEP-C, SCZ-T, and ANT-S subscales), these were not specific to the MND group; as shown in Table 2, while relatively few ADHD and No Diagnosis clients groups obtained high scores on the MAL, DEP-C and SCZ-T, the Clinical Controls and MND subjects both returned high scores on these scales. Further, both the ADHD and MND groups had equally high scores on the ANT-S subscale, significantly higher than the other two groups. As seen in Table 3, NIM was most successful of the PAI indices in identifying noncredible presentation at least when the proposed Musso et al. (2016) cut scores were applied.

Area under the curve (AUC) measures the overall performance of a given score classifier. Results from serial ROC analysis, including area under the curve (AUC) analyses, are included in Table 2. As may be seen, the only PAI variables that showed acceptable overall classification accuracy were NIM (AUC = 0.73) and two subscales: the SCZ-T subscale (AUC = 0.77) and DEP-C (AUC = 0.72).

Next, we examined the scores obtained by each group on the proposed PAI Scale-FAA. First, we removed those subjects whose scores on the PAI ICN ($n = 6$) or INF ($n = 6$)

Table 2. Means, *SD*, *F* value, effect size (Cohen's *d*), Area under the curve (AUC) and post hoc comparisons for validity scales of the PAI, proposed FAA scales and subscales, and proposed CAARS validity scales after removal of subjects with highly inconsistent or infrequent scores.

Scale	Mean/*SD*	ADHD (A)	Clinical control (CC)	MND	No diagnosis (ND)	*F* value	Post Hoc comparisons	Cohen's *d*	AUC
ICN	Mean	53.1	53.0	53.5	51.1	1.43		.31	.54
	SD	8.3	9.5	9.8	8.6				
INF	Mean	55.9	53.3	56.4	53.3	3.12*		.44	.57
	SD	8.2	7.9	8.6	8.0				
NIM	Mean	55.3	61.6	68.3	54.0	14.65**	ND,A < CC < MND	.99	.73
	SD	11.0	14.3	17.4	11.8				
PIM	Mean	43.3	39.4	39.8	45.8	5.68*	CC, MND < ND	.62	.41
	SD	9.5	11.7	12.5	11.6				
MAL	Mean	.66	1.08	1.43	.84	7.93**	A < CC,MND; ND < MND	.72	.66
	SD	.76	.95	1.12	.89				
RDF	Mean	−.586	−.313	−.461	−.701	2.04		.37	.53
	SD	1.0	1.19	1.0	1.04				
SCZ T	Mean	67.43	70.48	80.29	62.03	16.784**	ND < CC, MND; A < MND	1.06	.77
	SD	12.53	15.30	14.38	14.32				
DEP C	Mean	62.86	69.67	74.80	58.85	17.80**	ND, A < CC, MND	1.09	.72
	SD	11.37	15.42	15.11	13.16				
ANT S	Mean	61.15	54.79	61.29	54.62	6.40**	ND, CC < A, MND	.65	.57
	SD	13.17	12.80	16.36	12.91				
Scale-FAA	Mean	.475	.521	.585	.378	8.770**	ND < CC, MND; A < MND	.78	.65
	SD	.22	.26	.26	.25				
Item-FAA	Mean	.270	.300	.476	.212	4.59*	A, CC, ND < MND	.56	.64
	SD	.38	.35	.45	.32				
CII	Mean	12.2	10.4	15.0	10.8	2.264	CC, ND < MND	.62	.65
	SD	5.9	6.1	7.3	6.7				
Diss	Mean	11.9	12.6	24.1	12.4	14.685**	ND, CC, A < MND	1.61	.87
	SD	5.7	6.9	8.3	7.3				
EI	Mean	.70	.52	2.71	.67	7.99**	CC, ND, A < MND	1.19	.82
	SD	1.9	.8	2.3	1.4				

Note. INC: PAI Inconsistency scale; INF: PAI Infrequency scale; NIM: PAI Negative Impression scale; PIM: PAI Positive Impression scale; MAL: PAI Malingering scale; RDF: PAI Rogers' Discriminant Function; SCZ T: PAI Schizophrenia Thought Disorder; DEP C: PAI Depression Cognitive; ANT S: PAI Antisocial Features Stimulus Seeking; CII: CAARS Infrequency Index; DISS: E-Caars Dissimulation; EI: E-CAARS Exaggeration Index; MND: malingered neuropsychological dysfunction.
*$p < .05$;
**$p < .001$.

indicated that the protocols were answered in such an inconsistent or careless manner as to be invalid. As shown in Table 2, mean Scale-FAA score was significantly different across groups, $F_{(3,\ 315)} = 8.77$, $p < .001$. Tukey post hoc comparison showed the No Diagnosis group's mean Scale-FAA was significantly lower than both the Clinical Control group and the MND group, but the MND score was significantly higher than the ADHD group alone. As shown in Table 3, the percentage found over the cut score was significantly different among groups, $X^2_{(3)} = 14.6$, $p = .002$, with both the Clinical Control (adjusted residual $T = 2.2$) and the MND (adjusted residual $T = 1.6$) groups yielding significantly higher proportions and the No Diagnosis (adjusted residual $T = -2.2$) group yielding the lowest proportion. While 37.1% of the MND group scored above the recommended cutoff for this Scale-FAA, so too did 34.9% of the CC group and 22.4% of the ADHD group; only those in the no-diagnosis group were identified at a lower rate. The AUC scores for the Scale-FAA fell just below the acceptable range (see Table 2).

A slightly different picture emerged when evaluating the proposed Item-level FAA (Table 2 and 3). After removal of subjects with extreme ICN and INF, the mean Item-FAA score was different across groups, $F_{(3,\ 315)} = 4.59$, $p < .05$. Tukey post hoc comparisons showed that the MND group returned, on average, significantly higher scores than all other groups, however, the AUC score for this variable fell just below the acceptable range (.64). As shown in Table 3, the percentage found to be above the cut score was significantly different among groups, $X^2_{(3)} = 11.0$, $p < .012$ with the MND group being significantly larger (adjusted residuals $T = 2.7$) and the No diagnosis group being significantly smaller (adjusted residual $T = -2.1$). While the Item-FAA correctly identified 45.3% of MND individuals, Table 3 shows that a substantial number of students in the other groups also achieved a score above the recommended cutoff.

To examine the efficacy of other previously published self-report validity measures in identifying feigned ADHD, we next examined how well the CII and the E-CAARS measures functioned in identifying the MND group. Given that the CAARS technical manual recommends not interpreting protocols where the Inconsistency Index score is 8 or more, we first removed all such protocols from the dataset. This resulted in 174 clients. As noted above, not all students in this sample had been given the E-CAARS, and some records in the archival dataset included only scale as opposed to item-level responses for the CAARS. As such, after removal of invalid protocols we had a relatively smaller sample for these investigations ($n = 130$ for CII and $n = 126$ for E-CAARS, with some overlap).

As shown in Table 2, the four groups did not differ significantly in their CII scores, although the MND group did achieve higher mean scores than both the CC and ND groups. The AUC score (.65) fell below the acceptable classification range. In addition, Table 3 shows that the CII cut-score proposed by Suhr et al. (2011) identified relatively few MND individuals. By contrast, both the Dissimulation and EI scores from the E-CAARS identified a significantly larger

Table 3. Percentage per diagnostic group flagged as noncredible using PAI cut scores as suggested by Musso et al. (2016) and by CAARS cut scores as suggested by Suhr et al. (2011) and Harrison & Armstrong (2016).

PAI validity scale	Musso et al. cut scores	ADHD	Clinical control	MND	No diagnosis	X^2	*p*
ICN	% $T >= 70$	2.7	6.0	8.3	2.6	3.61	.307
INF	% $T >= 75$	2.7	0	2.8	1.7	1.93	.588
NIM	% $T >=77$	5.4	16.4	27.8	6.8	18.55	<.001
PIM	% $T <=25$	1.8	11.9	13.9	5.1	11.13	.011
MAL	% $>= 3$	0.9	9.1	17.1	5.1	14.58	.002
RDF	% $>= 1$	7.2	14.9	8.3	6.8	4.07	.254
PAI validity scales	Aita et al. cut scores						
Scale-FAA	% $>= 0.7$	22.4	34.9	37.1	13.2	14.6	.002
Item-FAA	% $>= 0.6$	24.3	27.0	45.5	17.5	11.0	.012
CAARS validity scale	Suhr/Harrison cut scores						
CII	% $>= 21$	6.8	9.7	17.6	5.3	2.57	.463
Dissimulation	% $>= 21$	7.0	10.0	76.5	11.1	44.88	<.001
EI	% $>=3$	9.3	3.4	52.9	8.3	26.46	<.001

Note. PAI: Personality Assessment Inventory; INC: Inconsistency; INF: Infrequency; NIM: Negative Impression Management; PIM: positive impression management; MAL: malingering index; RDF: Rogers Discriminant Function; MND: malingered neuropsychological dysfunction. Musso: cut scores recommended by Musso et al. (2016). Scale-FAA and Item-FAA as described by Aita et al. (2018) using PAI scores with invalid inconsistency and infrequency scores removed. CAARS: Conners' Adult Attention Rating Scale with invalid inconsistency scores removed; CII: CAARS infrequency Index (Suhr et al., 2011); Dissimulation: E-CAARS Dissimulation Scale; EI: E-CAARS Exaggeration Index (Harrison & Armstrong, 2016).

number of MND students than in any other group (Table 2), and both had excellent AUC classification ability. Furthermore, as shown in Table 3, 76.5% of the MND group were correctly classified using the DISS score, and 52.9% using the EI.

Performance characteristics for all variables of interest are shown in Table 4. The specificity of all the PAI-related validity scales is greater than .90 (the level suggested to ensure a low false positive rate; Boone, 2007; Larrabee, 2012, 2014; Vickery et al., 2001), while the FAA scale and item-based measures fall below .90; the FAA item measure had a specificity of .78, and the Scale-FAA was .79. Sensitivity of all PAI measures and the FAA scales was lower; the Item-FAA was highest at .45.

Evaluating performance of the alternative SVTs, the CII had acceptable specificity (see Table 4), but sensitivity to MND was weak (.18). Similarly, PPV was, at best, .52 and similar to values returned by most of the PAI measures. By contrast, the E-CAARS measures (Dissimulation and EI) demonstrated better psychometric performance. Specificity was above .90 and sensitivity was .76 and .53 for these two scores, respectively.

Given that the base rate of malingering ADHD is not known, Table 4 also provides clinicians with PPV and NPV scores for exaggeration base rates of 20% and 30%, along with the base rate of MND in the current study (12% for PAI data, 7% for CAARS data). Apart from the NIM which had acceptable PPV, other existing PAI validity scales showed weak predictive validity. Further, in Table 4 it may be seen that the PPV for the Scale-FAA is quite low. The probability of a person with a score above the recommended cutoff being an actual malingerer is, at best, 30% if using the scale-based FAA at a base rate of 30% malingerers. The Item-FAA fared somewhat better; at a 30% base rate this score had a 52% chance of accurately identifying a malingerer, a score comparable to the CII. For other PAI validity scales, the NIM had a PPV of 58% at a base rate of 30%, and MAL had a PPV of 59% if the base rate of malingering is 30%; for most other PAI indices PPV was approximately 50% or lower. Regarding NPV, a score below those recommended by Musso et al. (2016) for existing PAI validity scales has a high likelihood of correctly classifying honest individuals, except at higher base rates of malingering.

By contrast, PPV and NPV scores were strongest for the E-CAARS measures; PPV ranged from .57 to .78 for Dissimulation and .53 to .75 for EI, depending on base rate, and negative predictive values were also strong for these E-CAARS measures, except at higher estimated malingering base rates.

Discussion

Unlike most other studies evaluating exaggerated ADHD identification, the present study employed an ecologically valid sample of definite malingerers to investigate the predictive capabilities of two newly described SVTs derived from the PAI and compared these with existing SVT methods for identifying noncredible ADHD symptoms. Using an archival sample of postsecondary students who had all sought an assessment for possible ADHD, we compared the scores of those who met the revised Slick et al. (1999) criteria for definite malingered neurocognitive dysfunction with all other clinical groups (ADHD diagnosis, mental health diagnosis, and no diagnosis). Results demonstrated that neither of the newly described PAI algorithms was able to adequately detect those individuals classified as noncredible ADHD in this sample. Specificity of both algorithms failed to meet or exceed the .90 threshold recommended for minimal false positive diagnoses. Both in the current study and at differing estimated base rates, neither the scale- nor item-based FAA scores was able to correctly identify more than about 50% of those believed to be feigning ADHD, although the Item-FAA produced PPV results comparable to the CII. The limited utility of these algorithms was seen most clearly in the AUC analyses, demonstrating that both item- and scale-FAA scores have unsatisfactory classification rates. While students strongly suspected of definite malingered neurocognitive dysfunction based on multimodal validity criteria returned, on average, higher Scale- and Item-FAA scores than other groups, the cut-scores for these indices

Table 4. Sensitivity, specificity, PPV & NPV of PAI variables either determined by Aita et al. (2018) algorithms or by Musso et al. (2016) cut value and of the CAARS variables as defined by Suhr et al. (2011) and Harrison & Armstrong (2014).

		Current study base rate = 12%				20%		30%	
Test	Cut	Sen	Spec	PPV	NPV	PPV	NPV	PPV	NPV
PAI Item-FAA	0.6	.45	.78	.19	.92	.39	.85	.52	.77
PAI Scale-FAA	0.7	.36	.79	.17	.91	.16	.91	.30	.83
PAI ICN	70	.08	.97	.23	.90	.38	.81	.51	.70
PAI INF	75	.03	.98	.17	.89	.29	.80	.42	.70
PAI NIM	77	.28	.92	.29	.90	.45	.84	.58	.72
PAI PIM	25	.14	.95	.24	.90	.39	.81	.52	.71
PAI MAL	3	.32	.91	.17	.96	.46	.84	.59	.73
RDF	1	.08	.91	.10	.89	.19	.80	.29	.69
CAARS DI	21	.76	.91	.57	.96	.68	.94	.78	.81
EI	3	.53	.93	.53	.93	.64	.89	.75	.77
CII	21	.18	.93	.27	.88	.38	.82	.52	.71

Note. PAI: Personality Assessment Inventory; Item-FAA: Feigned Adult ADHD index using PAI items; Scale-FAA: Feigned Adult ADHD index using PAI Scales; ICN: Inconsistency scale; IFN: Infrequency Scale; NIM: Negative Impression Scale; PIM: Positive Impression Scale MAL: Malingering Scale; RDG: Rogers Discriminant Function; eCAARS DI: experimental versions of Conners Adult ADHD Rating Scale Dissimulation Score; EI: experimental versions of Conners Adult ADHD Rating Scale Exaggeration Index; CII: CAARS Infrequency Index; Sen: sensitivity; Spec: specificity; PPV: positive predictive value; NPV: negative predictive value.

failed to adequately identify those suspected of feigning; additionally, a sizable number of honest students would be falsely accused when using these algorithms.

With the exception of the NIM (and the SCZ-T and DEP-C subscale scores, which are not validity scales), other existing validity scales from the PAI had similar problems; they all had acceptable specificity but weaker sensitivity to feigned ADHD. The NIM scale alone had acceptable ability to identify noncredible performance in these students, which is consistent with results from previous studies examining use of the NIM at lower cutoff thresholds than recommended by the test manual (e.g., Morey, 2019; Rios & Morey, 2013; Smith, Cox, Mowle, & Edens, 2017). Given that Aita et al. (2018) specifically did not include NIM in their Scale-FAA calculation (as it was employed to identify potential malingerers in their ADHD sample), it seems that perhaps their FAA regression scale suffered from the lack of inclusion of this scale. The SCZ-T and DEP-C subscales did show good ability to identify noncredible presentation, but the performance characteristics of other PAI subscales included in the Scale-FAA (i.e., PIM, ANT-S) were not as strong. The NIM detects exaggerated unfavorable presentation based on bizarre and unlikely symptoms, and results from the current study suggest that this scale does indeed help to correctly identify individuals who are presenting in such a manner.

Employing the cut-scores in the present study that were recommended by Musso et al. (2016), the best that any of the existing PAI validity scales could achieve was a 59% probability that a score above the cut-score correctly identified a malingerer, and that was the MAL score at an estimated malingering base rate of 30%. Similarly, the CII derived from the CAARS had acceptable specificity but weak sensitivity, and PPV scores ranged from 27% in the current study to 52% at a base rate of 30% malingerers. In other words, at best there is about a 52% probability that someone with a score above 21 on the CII is truly feigning ADHD.

By contrast, the Dissimulation score and the EI derived from the E-CAARS showed stronger classification statistics. Both indices demonstrated acceptable specificity and had the highest levels of sensitivity as well. Further, ability to positively predict noncredible performance ranged from just over half to three-quarters of clinical feigners, depending on underlying base rate of feigning. The AUC analysis also demonstrated clearly that these two methods generated the best classification statistics in terms of overall accuracy.

Until now, most studies investigating ways to identify malingered ADHD have relied on simulation designs. Results from the present study demonstrate that analog malingerers may not behave similarly to real-world malingerers, and underscore the need to apply diagnostic formulae derived from analog samples to real-world cases in order to evaluate the ecological validity of such measures. Simulation designs, while having good internal validity, are often criticized because one cannot know their generalizability to actual, real-world malingerers (Haines & Norris, 1995; Rogers, Bagby, & Dickens, 1992). Generalizability is called into question because non-disabled simulators do not have the same motivation to malinger compared to actual malingers. Without such motivation, and given only immediate coaching or access to stereotypical symptoms, analog malingerers may over-estimate the deficits associated with specific disabilities (Haines & Norris, 1995) making them easy to identify relative to honest but disabled clinical subjects.

Harrison, Lee and Suhr (in press) conducted an extensive literature review of research into methods to detect feigned LD and ADHD. When comparing the test performance and symptom report of real-world clinical subjects who failed validity measures to those who passed, many studies find these groups are difficult to distinguish. Individuals classified in the real-world malingering groups endorse a greater number of symptoms than do individuals in the credible group, although frequently still within a believable range. A similar pattern of outcomes can be found for performance on neuropsychological tests. These findings suggest that individuals who fail PVT or who meet the revised Slick et al. (1999) criteria for malingering are often difficult to detect based solely on self-report or neuropsychological test performance. By contrast, Harrison et al. (in press) conclude that the vast majority of studies employing analog malingerers find significant differences between those feigning ADHD and individuals given these diagnoses. It is therefore unclear how well research from analog malingers generalizes to noncredible performers in actual clinical samples.

Findings from the present study highlight the continued need to investigate methods to identify accurately those feigning ADHD, especially in post-secondary populations. Almost all of the methods investigated at present fail to identify the majority of those strongly suspected of feigning; this may be due to the fact that, unlike other medical conditions, diagnosis of ADHD may be made based on a client demonstrating only six of a total of 18 symptoms of inattention or six of 18 symptoms of hyperactivity/impulsivity

(APA, 2013). Given that there can be many different permutations and combinations of exaggerated symptoms that could lead to successful deceit, those feigning may be able to take many different routes in order to successfully exaggerate symptoms of ADHD. As a result, clinicians likely need multiple methods to detect noncredible symptom reporting, primarily because there is no one consistent set of symptoms present even in those who genuinely have the condition.

This study has one significant limitation; clients were identified as probable feigners based on the revised Slick et al. (1999) criteria outlined by Slick & Sherman (2012) using performance validity tests. This poses two difficulties. First, one cannot know for certain whether these suspected individuals were truly malingerers, which in turn poses a threat to internal validity; some false positives may be included in this study, therefore diluting the sample. Recent research, however, suggests that the probability of false positive diagnosis of malingering is reduced to almost zero if a clinical subject fails more than one stand-alone validity measure during a psychological assessment (Jasinski, Berry, Shandera, & Clark, 2011; Musso & Gouvier, 2014; Sollman et al., 2010; Tucha, Fuermaier, Koerts, Groen, & Thome, 2015; Williamson et al., 2014). As such, and given that we sampled only from cases where the client came for assessment of possible ADHD, we believe that there is a high likelihood that those in the MND group were truly presenting noncredibly.

Second, it is also possible that this method of identifying the MND group failed to accurately identify all individuals who were feigning symptoms of ADHD. There is currently no "gold standard" for identifying feigned ADHD, and there are almost no stand-alone SVTs validated for use in ADHD assessments (Harrison et al, in press). As such, we classified noncredible performers based on multidimensional criteria that included PVT failure. The difficulty, however, is that there may be limited convergence between those who feign on performance measures and those who overreport symptoms, especially since ADHD is typically diagnosed based on self-reported symptoms rather than performance. In the ADHD literature in general there is little cross-method convergence between neuropsychological performance and self-reported difficulties (Barkley, 2019).

Research support for a higher-order dimension of generalized feigning is reportedly weak (Merton & Merckelbach, 2013) suggesting that the overlap between classification using PVTs and SVTs may be limited. For instance, Copeland et al. (2016) examined the differential relationship of PVTs and SVTs with cognitive performance and self-reported symptom measures. While not directly evaluating ADHD, they found that individuals who passed both a PVT and an SVT, as well as those who pass only a PVT, had better cognitive performance and self-reported fewer symptoms relative to those who failed both tests. They suggest that this finding supports the differential utility of PVTs and SVTs when assessing cognitive as opposed to self-reported symptoms. Similarly, Haggerty, Frazier, Busch, & Naugle (2007) found only modest or slight correlations between scores on the VSVT (a performance measure) and feigning measures derived from the PAI (self-report), implying that, in general, individuals feigning psychiatric impairment do not always underperform on cognitive measures of exaggeration. Finally, Dandachi-FitzGerald & Merckelbach (2013) found that a SVT examining general malingered symptoms outperformed the MSVT in correctly identifying analog malingerers, suggesting that memory oriented PVTs may be useful in detecting malingered memory problems but not other types of feigned psychopathology.

By contrast, some research (e.g., Green et al., 2001) suggests that failure on a PVT presumably assessing one cognitive domain is associated with lower scores on neuropsychological tests in both the same and different areas of functioning. In other words, poor performance on a PVT predicts generalized poor performance on all other neuropsychological tests administered. These findings would suggest that individuals feigning cognitive deficits do so across multiple cognitive domains in a somewhat indiscriminate manner.

There is research to support this view. For example, Lange, Iverson, Brooks, & Rennison (2010) found that while individuals evaluated for mild traumatic brain injury who failed the TOMM performed more poorly on cognitive measures than did those who passed, the failing individuals were also more likely to fail two other SVTs. Likewise, Lippa et al. (2014) found that individuals who failed the WMT were more likely to also fail a SVT such as the Fake Bad Scale (FBS; Lees-Haley, English, & Glenn, 1991). Finally, in the only study examining classification of feigned ADHD using two different validity methods, Shura, Denning, Miskey, & Rowland (2017) compared individuals classified as feigning using the TOMM (trial 1) versus those identified using various validity scales from the MMPI-2 RF (Ben-Porath & Tellegen, 2011). While they found that significantly more of their clinical sample failed a self-report validity measure from the MMPI-2 RF than the TOMM (44.7% vs. 19.3%), those who overreported symptoms on the MMPI-2 RF did not endorse significantly higher levels of ADHD symptoms on self-report than did credible subjects. By contrast, those who failed the PVT reported significantly higher ADHD symptoms than did those who passed an SVT. In other words, as has been suggested by Marshall et al. (2010), using PVTs that appear to include measures of attention may work well to also identify individuals feigning ADHD symptoms.

Nevertheless, because we cannot be sure that our classification method truly captured all those who were exaggerating ADHD symptoms, further research with other clinical populations needs to be done to examine the effectiveness of the FAA algorithms before they can be declared unhelpful.

Like most clinical research, this study has a number of other minor limitations. First, there were more men in the ADHD group than the other comparator groups. While demographic studies consistently identify that more males than females are diagnosed with ADHD (Visser et al., 2014), our findings should be replicated with groups comprised of equal numbers of men and women. Second, this study dealt exclusively with postsecondary students approaching an

assessment center for evaluation of ADHD. Our findings may not generalize to other populations.

Third, one might suggest that the E-CAARS fared well relative to the other indices because the CAARS data employed in the present study overlapped that used to create these scales. This is not, however, correct. In fact, the current archival sample included only 16.1% of the same individuals whose CAARS scores were examined in the development of the Dissimulation and EI scales. In other words, for the most part these two E-CAARS scales were applied to CAARS protocols not used to create the original scales. Finally, due to the fact that the E-CAARS is an experimental test, that not all cases included raw data for the CAARS responses, and after removal of inconsistent self-reports, the sample used to evaluate the E-CAARS and the CII were smaller than for the PAI. Research should replicate our investigation using a larger sample.

In conclusion, results from this study indicate that, apart from the NIM, the other validity scales of the PAI as well as the proposed FAA algorithms derived from this test have lower classification accuracy and provide minimal information to clinicians about the possibility of exaggerated symptom reporting. Some subscales of the PAI, however, merit further investigation of their potential utility to identify noncredible ADHD symptom reporting. Other methods derived from self-report ADHD scales appear more sensitive to such exaggeration, at least when classified using the Slick & Sherman (2012) criteria.

To improve diagnostic practice and reduce inappropriate access to accommodations and supports, researchers agree that clinicians need to administer performance and/or symptom validity tests when conducting such evaluations (Boone, 2017; Harrison, 2017; Ziegler & Boone, 2013). Given that most validity tests were developed using analog malingerers, findings from this study emphasize the need to validate the performance of such analog-derived measures in real-world situations in order to improve diagnostic accuracy. Clinicians should employ at least one well validated PVT and at least one SVT when evaluating credibility of complaints when undertaking ADHD assessments in young adults, so as to capture both symptom overreporting as well as noncredible performance. This recommendation is in line with advice given recently by other researchers (e.g., Musso & Gouvier, 2014; Tucha et al., 2015). The main difficulty, however, is finding ADHD-specific SVTs that have been evaluated using both analog and clinical populations. Findings from the present study indicate that the NIM from the PAI and the CAARS EI and DI may provide assessors with additional information regarding noncredible symptom reporting, and should be used along with existing stand alone PVTs in adult ADHD evaluations.

References

Abrams, S. J. (2005). Unflagged SATs. *Education Next*, *3*(Summer), 42–44.

Advokat, C. D., Guidry, D., & Martino, L. (2008). Licit and illicit use of medications for attention-deficit hyperactivity disorder in undergraduate college students. *Journal of American College Health*, *56*(6), 601–606. doi:10.3200/JACH.56.6.601-606

Aita, S. L., Sofko, C. A., Hill, B. D., Musso, M. W., & Boettcher, A. C. (2018). Utility of the Personality Assessment Inventory in detecting feigned Attention-Deficit/Hyperactivity Disorder (ADHD): The Feigned Adult ADHD index. *Archives of Clinical Neuropsychology*, *33*(7), 832–844. doi:10.1093/arclin/acx113

Alexander, S., & Harrison, A. G. (2013). Cognitive responses to stress, depression, and anxiety and their relationship to ADHD symptoms in first year. *Journal of Attention Disorders*, *17*(1), 29–37. doi:10.1177/1087054711413071

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (DSM-IV-TR)* (4th Ed). Washington, DC: American Psychiatric Association.

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

An, K., Zakzanis, K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology*, *27*(8), 849–857. doi:10.1093/arclin/acs085

Barkley, R. (2019). Neuropsychological testing is not useful in the diagnosis of ADHD: Stop it (or prove it)! *The ADHD Report*, *27*, 1–7.

Belkin, D. (2018, May). Colleges bend the rules for more students, Give them extra help. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/colleges-bend-the-rules-for-more-students-give-them-extra-help-1527154200.

Ben-Porath, Y. S., & Tellegen, A. (2011). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form): Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.

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

Boone, K. B. (2007). *Assessment of feigned cognitive impairment. A neuropsychological perspective*. New York, NY: Guilford.

Boone, K. B. (2017). *Neuropsychological evaluation of somatoform and other functional somatic conditions: Assessment primer*. New York, NY: Taylor and Francis.

Conners, C. K., Erhardt, D., & Sparrow, E. (1998). *Conners' Adult ADHD Rating Scales (CAARS): ADHD across the life span*. Toronto, ON: Multi-Health Systems.

Cook, C., Bolinger, E., & Suhr, J. A. (2016). Further validation of the CAARS Infrequency Index (CII) for detection of noncredible report of ADHD symptoms. *Archives of Clinical Neuropsychology*, *31*(4), 358–364. doi:10.1093/arclin/acw015

Copeland, C. T., Mahoney, J. J., Block, C. K., Linck, J. F., Pastorek, N. J., Miller, B. I., … Sim, A. H. (2016). Relative utility of performance and symptom validity tests. *Archives of Clinical Neuropsychology*, *31*(1), 18–22. doi:10.1093/arclin/acv065

Dandachi-FitzGerald, B., & Merckelbach, H. (2013). Feigning ≠ feigning a memory deficit: The Medical Symptom Validity Test as an example. *Journal of Experimental Psychopathology*, *4*(1), 46–63. doi:10.5127/jep.025511

Danielson, M. L., Bitsko, R. H., Ghandour, R. M., Holbrook, J. R., Kogan, J. D., & Blumberg, S. J. (2018). Prevalence of parent-reported ADHD diagnosis and associated treatment among U.S. children and adolescents, 2016. *Journal of Clinical Child & Adolescent Psychology*, *47*(2), 199–212. doi:10.1080/15374416.2017.1417860

Frazier, T.,W., Frazier, A., R., Busch, R.,M., Kerwood, M., A., & Demaree, H., A. (2008). Detection of simulated ADHD and reading disorder using symptom validity measures. *Archives of Clinical Neuropsychology*, *23*(5), 501–509. doi:10.1016/j.acn.2008.04.001

Green, P. (2003). *Word memory test for windows: User's manual and program*. Edmonton: Green's Publishing.

Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows. User's manual*. Edmonton, Canada: Green''s Publishing.

Green, P. (2008). *Advanced interpretation program*. Edmonton, Canada: Green's Publishing.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claims. *Brain Injury*, *15*(12), 1045–1060. doi:10.1080/02699050110088254

Greenberg, L. M., Kindschi, C. L., Dupuy, T. R., & Hughes, S. J. (2007). *T.O.V.A. clinical manual: Test of Variables of Attention continuous performance test*. Los Alamitos, CA: The TOVA Company.

Greenfield, B., Ochs, E., & Hechtman, L. (2002). Diagnosing attention-deficit/hyperactivity disorder in adults: Challenges and controversies. *Drug Benefit Trends*, *14*(4), 27–33. Retrieved from https://psycnet.apa.org/record/2002-13156-002

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

Haggerty, K., Frazier, T., Busch, R., & Naugle, R. (2007). Relationships among Victoria Symptom Validity Test indices and personality assessment inventory validity scales in a large clinical sample. *The Clinical Neuropsychologist*, *21*(6), 917–928. doi:10.1080/13854040600899724

Haines, M. E., & Norris, M. P. (1995). Detecting the malingering of cognitive deficits: An update. *Neuropsychology Review*, *5*(2), 125–148. doi:10.1007/BF02208438

Hanson, C. L., Burton, S. H., Giraud-Carrier, C., West, J. H., Barnes, M. D., & Hansen, B. (2013). Tweaking and tweeting: Exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. *Journal of Medical Internet Research*, *15*(4), e62. doi:10.2196/jmir.2503

Harrison, A. G. (2004). An investigation of reported symptoms of ADHD in a university population. *The Adhd Report*, *12*(6), 8–11. doi:10.1521/adhd.12.6.8.55256

Harrison, A. G. (2017). Clinical, ethical, and forensic implications of a flexible threshold for LD and ADHD in postsecondary settings. *Psychological Injury and Law*, *10*(2), 138–150. doi:10.1007/s1220

Harrison, A. G., Alexander, S., & Armstrong, I. A. (2013). Higher reported levels of depression, stress, & anxiety are associated with increased endorsement of ADHD symptoms by postsecondary students. *Canadian Journal of School Psychology*, *28*(3), 243–260. doi:10.1177/0829573513480616

Harrison, A. G., & Armstrong, I. (2014). WISC-IV unusual digit span performance in a sample of adolescents with learning disabilities. *Applied Neuropsychology: Child*, *3*(2), 152–160. doi:10.1080/21622965.2012.753570

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

Harrison, A. G., & Edwards, M. J. (2010). Symptom exaggeration in post-secondary students: Preliminary base rates in a Canadian sample. *Applied Neuropsychology*, *17*(2), 135–143. doi:10.1080/09084281003715642

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

Harrison, A. G., Lee, G., & Suhr, J. (in press). Use of performance validity tests and symptom validity tests in assessment of specific learning disorders and attention deficit hyperactivity disorder. In Kyle Boone (Ed.). *Assessment of feigned cognitive impairment* (2nd ed.). New York: Guilford.

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

Jasinski, L. J., Berry, D. T. R., Shandera, A. L., & Clark, J. A. (2011). Use of the Wechsler Adult Intelligence Scale Digit Span subtest for malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology*, *33*(3), 300–314. doi:10.1080/13803395.2010.516743

Kessler, R. C., Adler, L., Barkley, R., Biederman, J., Conners, C. K., Demler, O., … Zaslavsky, A. M. (2006). The prevalence and correlates of adult ADHD in the United States: Results from the National Comorbidity Survey replication. The *American Journal of Psychiatry*, *163*(4), 716–723. doi:10.1176/ajp.2006.163.4.716

Kirk, J. W., Hutaff-Lee, C. F., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The relationship between the self-report BASC-2 validity indicators and performance validity test failure after pediatric mild traumatic brain injury. Assessment, *21*(5), 562–569. doi:10.1177/1073191114520626

Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The value of the WISC-IV Digit Span Subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology*, *26*(5), 377–384. doi:10.1093/arclin/acr040

Lange, R. T., Iverson, G. L., Brooks, B. L., & Rennison, V. L. (2010). Influence of poor effort on self-reported symptoms and neurocognitive test performance following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *32*(9), 961–972. doi:10.1080/13803391003645657

Lange, R.T., & Lippa, S. (2017). Sensitivity and specificity should never be interpreted in isolation without consideration of other clinical utility metrics. *The Clinical Neuropsychologist*, *31*(6–7), 1015–1028. doi:10.1080/13854046.2017.1335438

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, *22*(4), 666–679. doi:10.1080/13854040701494987

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*(4), 625–630. doi:10.1017/S1355617712000240

Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology*, *29*(4), 364–373. doi:10.1093/arclin/acu019

Leark, R., Greenberg, L., Kindschi, C., Dupuy, T., & Hughes, S. (2008). *T.O.V.A. professional manual.* Los Alamitos, CA: The TOVA Company.

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A fake bad scale on the MMPI-2 for personal injury claimants. *Psychological Reports*, *68*(1), 203–210. doi:10.2466/pr0.1991.68.1.203

Lewak, D. (2018, May). Rich parents are using doctor's notes to help kids cheat the SATs. *New York Post*. Retrieved from https://nypost.com/2018/05/02/rich-parents-are-using-doctors-notes-to-help-kids-cheat-the-sats/

Lippa, S. M., Pastorek, N. J., Romesser, J., Linck, J., Sim, A. H., Wisdom, N. M., & Miller, B. I. (2014). Ecological validity of performance validity testing. *Archives of Clinical Neuropsychology*, *29*(3), 236–244. doi:10.1093/arclin/acu002

Mannuzza, S., Klein, R. G., Klein, D. F., Bessler, A., & Shrout, P. (2002). Accuracy of adult recall of childhood attention deficit hyperactivity disorder. *American Journal of Psychiatry*, *159*(11), 1882–1888. doi:10.1176/appi.ajp.159.11.1882

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. doi:10.1037/pas0000247

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., & Barker, J. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist*, *24*(7), 1204–1237. doi:10.1080/13854046.2010.514290

McCabe, S. E., Teter, C. J., & Boyd, C. J. (2006). Medical use, illicit use and diversion of prescription stimulant medication. *Journal of Psychoactive Drugs*, *38*(1), 43–56. doi:10.1080/02791072.2006.10399827

McGough, J. J., & Barkley, R. A. (2004). Diagnostic controversies in adult attention deficit hyperactivity disorder. *American Journal of Psychiatry*, *161*(11), 1948–1956. doi:10.1176/appi.ajp.161.11.1948

Merton, T., & Merckelbach, H. (2013). Introduction to malingering research and symptom validity assessment. *Journal of Experimental Psychopathology*, *4*, 3–5. doi:10.1177/2043808713004001O2

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2012). Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychology*, *27*(1), 10–22. doi:10.1093/arclin/acr084

Mitchell, H. (2012, January 25). Faking ADHD gets you into Harvard. *The Daily Beast*. Retrieved from http://www.thedailybeast.com/articles/2012/01/25/faking-adhd-gets-you-into-harvard.html

Morey, L. C. (1991). *The Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (1993). Defensiveness and malingering indices for the PAI. Paper presented at the meeting of the American Psychological Association, Toronto, Canada.

Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory*. Odessa, FL: Psychological Assessment

Morey, L. C. (2019). Examining a novel performance validity task for the detection of feigned attentional problems. *Applied Neuropsychology: Adult*, *26*(3), 255–267. doi:10.1080/23279095.2017.1409749

Musso, M. W. (2013). *Detecting malingered ADHD using the personality assessment inventory: An exploratory analysis in college students* (LSU Doctoral Dissertations) 39. Louisiana State University, Baton Rouge, LA. Retrieved from https://digitalcommons.lsu.edu/gradschool_dissertations/39.

Musso, M. W., & Gouvier, W. D. (2014). Why is this so hard? A review of detection of malingered ADHD in college students. *Journal of Attention Disorders*, *18*(3), 186–201. doi:10.1177/1087054712441970

Musso, M. W., Hill, B. D., Barker, A. A., Pella, R. D., & Gouvier, W. D. (2016). Utility of the Personality Assessment Inventory for detecting malingered ADHD in college students. *Journal of Attention Disorders*, *20*(9), 763–774. doi:10.1177/1087054714548031

Odland, A. P., Lammy, A. B., Martin, P. K., Grote, C. L., & Mittenberg, W. (2015). Advanced administration and interpretation of multiple validity tests. *Psychological Injury and Law*, *8*(1), 46–63. doi:10.1007/s12207-015-9216-4

Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology*, *18*(4), 379–395. doi:10.1093/arclin/18.4.379

Rios, J., & Morey, L. C. (2013). Detecting feigned ADHD in later adolescence: An examination of three PAI–A Negative Distortion Indicators. *Journal of Personality Assessment*, *95*(6), 594–599. doi:10.1080/00223891.2013.821071

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *The SIRS test manual*. Tampa, FL. Psychological Assessment Resources.

Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment*, *67*(3), 629–640. doi:10.1207/s15327752jpa6703_15

Ruiz, M. A., Drake, E. B., Glass, A., Marcotte, D., & van Gorp, W. G. (2002). Trying to beat the system: Misuse of the Internet to assist in avoiding the detection of psychological symptom dissimulation. *Professional Psychology: Research and Practice*, *33*(3), 294–299. doi:10.1037/0735-7028.33.3.294

Shura, R. D., Denning, J. H., Miskey, H. M., & Rowland, J. A. (2017). Symptom and performance validity with veterans assessed for attention-deficit/hyperactivity disorder (ADHD). *Psychological Assessment*, *29*(12), 1458–1465. doi:10.1037/pas0000436

Slick, D. J., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *Manual for the Victoria Symptom Validity Test*. Odessa, FL: Psychological Assessment Resources.

Slick, D. J., & Sherman, E. (2012). Differential diagnosis of malingering and related clinical presentations. In E. M. S. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 113–135). New York, NY: Oxford University Press.

Slick, D. J., Sherman, E., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*(4), 545–561. doi:10.1076/1385-4046(199911)13:04;1-Y;FT545

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

Sollman, M. J., Ranseen, J. D., & Berry, D. T. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, *22*(2), 325–335. doi:10.1037/a0018857

Suhr, J., Hammers, D., Dobbins-Buckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Archives of Clinical Neuropsychology*, *23*(5), 521–530. doi:10.1016/j.acn.2008.05.003

Suhr, J.A., Buelow, M., & Riddle, T. (2011). Development of an infrequency index for the CAARS. *Journal of Psychoeducational Assessment*, *29*(2), 160–170.

Suhr, J. A., & Gunstad, J. (2007). Coaching and malingering: A review. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits*. New York, NY: Oxford University Press.

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology*, *14*(3), 189–207. doi:10.1080/09084280701509083

Sussman, Z.W., Peterson, R., Connery, A., Baker, D., & Kirkwood, M. (2017). Utility of matrix reasoning as an embedded performance validity indicator in pediatric mild traumatic brain injury. *Applied Neuropsychology: Child*, *23*, 1–6. doi:10.1080/21622965.2017.1382359

Tapper, J., Morris, D., & Setrakian, L. (2006, March 30). Does loophole give rich kids more time on SAT? *ABC News*. Retrieved from http://abcnews.go.com/Nightline/story?id=1787712

Taylor, K. (2019, March 13). Fallout from the U.S. College admissions scandal: Arrests, damage control, and a scramble for answers. *The New York Times*.

Tombaugh, T. N. (1996). *TOMM: Test of memory malingering*. North Tonawanda, NY: Multi-Health Systems.

Tucha, L., Fuermaier, A., Koerts, J., Groen, Y., & Thome, J. (2015). Detection of feigned attention deficit hyperactivity disorder. *Journal of Neural Transmission*, *122*(S1), 123–S134. doi:10.1007/s00702-014-1274-3

Vickery, C. D., Berry, D. T. R., Inman, T., Harris, M., & Orey, S. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, *16*, 45–73. doi:10.1093/arclin/16.1.45

Visser, S. N., Danielson, M. L., Bitsko, R. H., Holbrook, J. R., Kogan, M. D., Ghandour, R. M., … Blumberg, S. J. (2014). Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003–2011. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(1), 34–46.e2. doi:10.1016/j.jaac.2013.09.001

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. *The Clinical Neuropsychologist*, *28*(7), 1182–1196. doi:10.1080/13854046.2014.956674\

Ziegler, E. A., & Boone, K. B. (2013). Symptom invalidity on neuropsychological testing. In P. Arnett (Ed.). *Secondary influences on neuropsychological test performance*. New York, NY: Oxford University Press.