August 2018

# The Detection of ADHD and Malingering in Young Adults

Heather Potts
*Syracuse University*

Abstract

Over the past 20 years, an increasing number of adults and college students have self-referred for ADHD evaluations. With the rise in adult ADHD evaluations, there has been increased concern that a proportion of these adults may be malingering the symptoms of ADHD to receive external incentives such as academic accommodations and stimulant medications. Research supports the use of well-validated measures to classify malingering in non-ADHD populations, yet all available validity tests have insufficient research to support their usage to detect this population. The present study investigated the ability of the Multidimensional ADHD Rating Scale (MARS) and two published validity tests (Word Memory Test and CAT-A Infrequency scale) to detect a group of non-ADHD college students instructed to feign ADHD, and to differentiate ADHD from non-ADHD cases. Results found that the MARS Symptom Validity Index demonstrated higher sensitivity rates for simulated malingering (75.4%) at close to optimal specificity (86.8%) compared to two published tests (sensitivity < 50%). The MARS Total Symptom index differentiated ADHD from non-ADHD cases with high sensitivity (87.1%). The study provides additional support for the effectiveness of the MARS symptom, impairment, and symptom validity indices to detect simulated cases of malingering, and to differentiate ADHD from non-ADHD cases.

Keywords: *ADHD, malingering, feigned ADHD, validity test, college students*

# THE DETECTION OF MALINGERING AND ADHD IN YOUNG ADULTS

by

Heather E. Potts

B.A., James Madison University, 2006
M.A., The George Washington University, 2008
M.S., Syracuse University, 2016

Dissertation
Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in *School Psychology*

Syracuse University
August 2018

Acknowledgements

This dissertation would not have been made possible without the guidance and support from my colleagues, family, and friends. First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Lawrence Lewandowski. I would not be the researcher or clinician I am without his assistance and guidance over the years, and for that I cannot thank him enough.

I would like to thank my dissertation committee, Drs. Kevin Antshel, Brian Martens, Shannon Sweeney, Leslie Gellis, and Catherine Engstrom, for their thoughtful advice and feedback on this project. I am incredibly thankful for the support I received from my internship supervisors, Dr. Anna Hickey and Erin Meyer, as I completed my graduation requirements. I must also recognize the Larry Lab members, Tim Reid, Kaya Wechsler, and Katie Herer, for their assistance with data collection.

I am deeply indebted to my mentors at James Madison University who fostered my passion for psychology and research, Dr. James Benedict and Dr. William Ernst. My pursuit of this doctoral degree would never have occurred without your early direction and continued encouragement to pursue graduate education.

To all my friends who have provided limitless support throughout this process, both from near and far, I give great thanks. Special thanks are made to all my JMU and Lo.Co. friends, in particular Amanda Flores, Megan Osborn, Ally Kelly, Kim Bell, Carol Spears, and Nichole Bernard. I am forever grateful to Megan Kremer and Shelly Blecha for the amazing memories and random snail mail cards of encouragement. Thank you to my brother's friends whom I am happy to call my brothers as well, Alex Burdick, Matt Harrington, and Chris Wolfe. The SU and Fellows Ave. crew provided much needed respite and support, and for that I extend major thanks to Laura Spenceley, Sean Pfiel, David Lemon, Oliver Haney, Lynaye Stone, Marlene Wong, Ryan Nichols, Daria Bakina, Whitney Wood, Jordan Wood, and Suzanne Spinola. Massive thanks go to Laura Spenceley and Whitney Wood who served double duty as both research advisor and best friend through the years. I am also very thankful for the amazing friendships I developed through the Teaching Mentoring program with Giovanna Urist, Tom Guiler, Tess Dussling, Bob Searing, and Catriona Stanfield.

Last but certainly not least, I must give my deepest and warmest thanks to my family who has provided me endless support throughout the years. I am eternally thankful for the help and constant encouragement I received from my parents, Allan and Lannie, to pursue my passion and to never give-up my dreams. Marla, I am so very lucky to have you as my sister and I owe you a great deal of gratitude for everything. My heart extends to my niece who sends me handmade cupcakes and the best artwork an aunt could ask for. To my nephew, I must say thank you for just simply being the best. To my brother and partner-in-crime throughout the lifespan, Tom, I am without words of thanks and hopefully one day we will live on the same coast again. My sister-in-law, Sarah, is thanked for her kindness and support. I thank my aunt, Helen Price, for her constant support and random visits. I must also express my gratitude to my late brother-in-law, Russ French, who opened my eyes to field of psychology and to amazing music in early life.

HP

TABLE OF CONTENTS

# The Detection of ADHD and Malingering in Young Adults

Since the 1990's, there has been increased clinical acceptance that Attention-Deficit/Hyperactivity Disorder (ADHD) can be diagnosed for the first time in adulthood (Barkley, Murphy, & Fischer, 2008; Davidson, 2008). Although the clinical community generally agrees that the diagnosis of ADHD can be conferred on adults, there is currently no agreement on best practices or standard diagnostic methods for diagnosing such cases (Harrison, 2017). With an increasing number of young adults self-referring for ADHD evaluations on college campuses (Barkley et al., 2008; Harrison, 2004; Weyandt & DuPaul, 2006), some have questioned whether current assessment practices are flawed, and in fact, are susceptible to false positive diagnoses (Harrison, 2017; Musso & Gouvier, 2014).

Individuals with ADHD are often provided with academic accommodations and stimulant medications, which non-ADHD college students report as positive incentives to have the diagnosis (Benson, Flory, Humphreys, & Lee, 2015; Jasinski & Ranseen, 2011; Lewandowski, Lambert, Lovett, Panahon, & Sytsma, 2014). Such incentives appear to be motivating a substantial minority of young adults to feign the disorder, with malingering occurring in approximately 15% to 50% of adult ADHD evaluations (Harrison & Edwards, 2010; Sullivan, May, & Galbally, 2007). Individuals who falsely receive the ADHD diagnosis could be provided with accommodations that could result in an academic advantage (Jasinski & Ranseen, 2011), as well as prescribed stimulants that could place them at increased medical risk (Benson et al., 2015; Park & Haning, 2016). Thus, there is a clear need to refine our diagnostic practices to accurately differentiate true cases of ADHD from those malingering the disorder (Musso & Gouvier, 2014).

Research indicates that standard ADHD assessments and practices have limited ability to detect malingering. With little preparation, motivated individuals can easily fake an ADHD

profile by under-performing on psychological/neuropsychological measures and over-reporting on rating scales (Marshall, Hoelzle, Heyerdahl, & Nelson, 2016). Although malingerers tend to exaggerate performance and responses, feigned test scores often still fall at levels that are comparable to a typical ADHD profile (Marshall et al., 2010; Tucha, Fuermaier, Koerts, Groen, & Thome, 2015). Because the malingering profile is not easily differentiated from ADHD on standard diagnostic measures (e.g., symptom rating scales and continuous performance tests), diagnosticians are at increased risk of rendering false ADHD diagnoses when relying upon these measures (Musso & Gouvier, 2014; Tucha et al., 2015). For these reasons, it is recommended that clinicians employ validity and effort tests, and for researchers to develop specialized tests that can reliably detect feigned ADHD (Musso & Gouvier, 2014; Sagar, Miller, & Erdodi, 2017).

Validity tests are designed to detect poor performance effort and/or exaggeration of symptoms associated with feigning (Rogers, 2008). Although many validity tests exist to detect feigned neurocognitive impairment and psychiatric disorders, research has found few validity tests that are moderately sensitive to the unique response bias of feigned ADHD (Musso & Gouvier, 2014; Tucha et al., 2015). This appears to be, in part, because the measures being used were not originally designed to detect malingering in the ADHD population. Higher detection of malingered ADHD appears to be feasible when a validity test is specifically designed to detect this population (Potts, 2016). While research has identified a few validity tests that could be effective, to date, none have sufficient evidence to support their usage to detect cases of feigned ADHD (Musso & Gouvier, 2014).

An ADHD diagnosis appears contingent upon the accuracy of available assessments to not only detect ADHD, but also to discriminate false report of malingering. The purposes of the present study were (a) to examine the ability of a new measure, the Multidimensional ADHD Rating Scale (MARS) to differentiate college students with ADHD diagnoses from simulated

malingering and non-ADHD controls, (b) to assess the accuracy of the MARS Symptom Validity Index (SV-index) to classify simulated malingering, and (c) to compare the classification accuracy of the MARS SV-index to that of the Word Memory Test (WMT) and the Clinical Assessment of Attention Deficit-Adult (CAT-A).

## Attention-Deficit/Hyperactivity Disorder

ADHD is classified as a neurodevelopmental disorder in the *Diagnostic and Statistical Manual for Mental Disorders-Fifth Edition* (*DSM-5*; American Psychiatric Association, 2013). ADHD is characterized by symptoms of inattention, hyperactivity, and impulsivity that started prior to the age of 12 (Criteria A and B). The symptom threshold for children is six or more inattention and/or six or more hyperactivity-impulsivity symptoms; but only five symptoms are needed to diagnose adults aged 17 or older. Symptoms must be documented in at least two settings (Criterion C) and symptoms must "interfere with or negatively affect" daily functioning and activities (Criterion D). Lastly, symptoms cannot be better explained by another disorder, such as substance use/abuse or mood disorders (Criterion E; American Psychiatric Association, 2013). See Appendix A for the full *DSM-5* criteria.

The ADHD diagnosis includes three different subtypes and additional severity specifiers to capture individual differences in composition of symptoms and impairment. Three symptom presentations of ADHD can occur based on the distribution of the 18 *DSM-5* symptoms. For individuals over the age of 17, the Predominantly Inattentive Presentation requires the presence of five or more inattentive symptoms occurring often or very often within the past six months, while the Predominantly Hyperactive-Impulsive Presentation requires with the presence of five or more hyperactive-impulsive symptoms (six symptoms per subtype are required for ages 16 and younger). The Combined Presentation is diagnosed if criteria for both the inattention and hyperactivity-impulsivity criteria are met. In addition to symptom presentations, severity

specifiers of *mild*, *moderate*, or *severe* can be used to characterize the frequency and intensity of symptoms and impairment. Lastly, i*n partial remission* can be used if an individual with a former diagnosis of ADHD no longer meets the symptom criteria, but continues to experience impairment of functioning (American Psychiatric Association, 2013).

ADHD is a disorder that can be diagnosed at any age and is presumed to be continuous throughout the lifespan in the majority of cases (Barkley et al., 2008). Although more frequently diagnosed in childhood (Centers for Disease Control and Prevention, 2014), ADHD can be diagnosed for the first time in adults (Barkley et al., 2008). A first time ADHD diagnosis in adulthood still requires onset of symptoms in childhood (≤ 12 years old; American Psychiatric Association, 2013). However, ADHD symptoms could go undiagnosed in childhood for a variety of reasons. Some parents may be hesitant to pursue a diagnostic evaluation for their child due to misconceptions associated with the ADHD label (Bussing, Zima, Mason, Porter, & Garvan, 2011; Pescosolido et al., 2008). Additionally, a child with ADHD symptoms may have functioned well, perhaps due to compensatory strategies, within structured and supportive settings (e.g., scheduled school day, parental monitoring). Without experiencing impairment in functioning, there may have been little reason or motivation for the child to receive a diagnostic evaluation (Barkley et al., 2008). However, some individuals may begin to experience impairment as they transition to adult settings that often require a greater degree of independence and executive functioning skills (e.g., college; Barkley et al., 2008). Thus, although ADHD is considered a lifelong disorder, an ADHD diagnostic evaluation may not be pursued until adulthood (Barkley et al., 2008; Lasky et al., 2016).

## Barriers to Accurate ADHD Diagnoses in Adults

There are many challenges to the accurate diagnosis of ADHD, especially first time diagnoses in adulthood. Across all ages, a diagnosis is complicated by a heterogeneous ADHD

symptom and cognitive profile (Barkley et al., 2008; Sonuga-Barke, 2005). ADHD encompasses individuals with different combinations and severity of symptoms that all meet the *DSM-5* diagnostic criteria. In addition to the listed *DSM-5* symptoms, ADHD is a neurobiological disorder associated with underlying deficits in executive functioning (EF) skills, including working memory, inhibitory control (i.e., self-regulation), and cognitive flexibility (i.e., shifting thinking between two concepts; Sonuga-Barke, 2005). Evidence also shows that individual differences in neurodevelopment and the environmental context can affect the type, timing, and expression of ADHD symptoms and neurocognitive skills across the lifespan (Barkley et al., 2008; Sonuga-Barke, 2005). Consequently, people with ADHD may report varying symptoms and neurocognitive skills at different age groups (Davidson, 2008). For example, a greater degree of inattention symptoms are reported by adults with ADHD compared to children with the disorder (Barkley et al., 2008; Davidson, 2008). The heterogeneity of the ADHD profile creates a wide range of ADHD phenotypes, and subsequently, a diverse range of clinical presentations that could meet criteria for the diagnosis.

Adult ADHD assessment is further limited by difficulties related to assessing the history of symptoms, obtaining collateral rater reports, and obtaining school/academic records. Although school records and parent reports may be relatively easy to obtain for children, the acquisition of collateral reports is more challenging with adult clients (Davidson, 2008). For example, adult clients have the right to refuse consent for the release of records without prejudicing their relationship with a therapist. Even if the adult client grants consent, parents/guardians and school records may be largely unavailable or unattainable. For these reasons, it is common for diagnostic decisions of adult ADHD to rely upon clinical interview and self-report measures (Davidson, 2008; Nelson, Whipple, Lindstrom, & Foels, 2014). Self-reports may be able to

effectively capture ADHD symptoms and impairment, but these methods are limited and susceptible to biased responding (Musso & Gouvier, 2014; Suhr, Cook, & Morgan, 2017).

One limitation of self-report measures is that subjective report of ADHD symptoms is not unique to individuals with an ADHD diagnosis. Self-report of ADHD symptoms is commonly reported by both children and adults *without* ADHD (Bird et al., 1988; DuPaul, Reid, Anastopoulos, & Power, 2014; Harrison, 2004; Lewandowski, Lovett, Codding, & Gordon, 2008). Specific to college campuses, Lewandowski and colleagues (2008) found that students without ADHD ($n$ = 496) endorsed an average of 4.5 out of 18 symptoms, indicating that most college students report ADHD symptoms to some degree (Lewandowski et al., 2008). As ADHD symptoms are commonly reported by the general college population (Lewandowski et al., 2008), a symptoms-only assessment may increase the risk of false diagnosis of non-ADHD individuals (Gathje, Lewandowski, & Gordon, 2008).

One way to reduce false positive diagnosis is the joint assessment of symptoms and impairment (Lewandowski et al., 2008). Although many individuals report symptoms, not all of them will also report impairment at sufficient levels to warrant a clinical diagnosis (Bird et al., 1988; DuPaul et al., 2014; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000; Wakefield, 2010). In illustration, DuPaul and colleagues (2014) had 1,070 teachers rate ADHD symptoms and impairment for two randomly selected student within their $6^{th}$-$12^{th}$ grade classrooms. Based upon the teachers' ratings, almost one in five students (18.9%) met the ADHD symptom count. This rate dropped to 7.3% when both symptoms and functional impairment were considered jointly, which is closer to accepted epidemiological base rates (DuPaul et al., 2014). Additional studies have documented a reduction in false positive diagnoses when one requires both high symptoms and impairment (Bird et al., 1988; DuPaul et al., 2014; Gathje et al., 2008; Shaffer et al., 2000). Thus, the assessment of impairment in ADHD evaluations is an important criterion

that seemingly increases accuracy of diagnostic decisions (Barkley et al., 2008; Gathje et al., 2008; Gordon et al., 2006).

Although assessment of symptoms and impairment appear to reduce false diagnosis of *honest* reporting individuals, high endorsement on symptoms and impairment rating scales alone does not automatically reflect a genuine case of ADHD. Rating scales are subjective measures and are easily biased or faked, particularly if one is motivated to do so (Marshall et al., 2016; Quinn, 2003). Therefore, high levels of symptoms and impairment could reflect either cases of clinical ADHD or exaggerated reports by non-ADHD individuals (e.g., Marshall et al., 2016). Thus, there is a need to develop methods that differentiate those who fake a clinical disorder from those who have the disorder.

## Motivation to Feign ADHD

Individuals could be motivated to feign a clinical disorder for a variety of reasons (Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998). In the *DSM-5*, malingering is not classified as a disorder, and is defined as the "intentional production of false or grossly exaggerated physical or psychological symptoms" with motivations to obtain external incentives (American Psychiatric Association, 2013, p. 726). Although an individual who is malingering is "faking," as noted by Iverson (2004), not all cases of "faking" are malingering. For example, some individuals may over-report symptoms unconsciously as a "cry for help," or as a way to obtain attention and help for their challenges (Iverson, 2006). Others may intentionally or unintentionally want a diagnosis to provide an explanation for their problems and failures (Rogers et al., 1998; Suhr & Wei, 2013), and still others may arrive to the evaluation with a belief that they have ADHD, and falsely over-report to confirm their beliefs (Barkley et al., 2008; Suhr & Wei, 2013). Yet, regardless of different motivations or incentives, all types of feigned reports and performances can affect diagnosis and treatment planning (Rogers, 2008).

While internal factors can motivate an individual to feign ADHD, the external incentives associated with the diagnosis increase the probability of malingering in college students (Jasinski & Ranseen, 2011; Musso & Gouvier, 2014). Individuals with a diagnosis of ADHD could be prescribed stimulants, be eligible for academic support services (e.g., tutoring, assistive technology) and qualify for test accommodations (e.g., extended time, private room testing) under *The Americans with Disabilities Act Amendments Act* (2008). Several studies have indicated that many college students view these incentives positively, and some are motivated to obtain them (Benson et al., 2015; Jasinski & Ranseen, 2011; Lewandowski et al., 2014). Stimulant medication is associated with greater concentration, inhibition, and memory in adults, which can result in a small to moderate advantage in academic contexts (Ilieva, Hook, & Farah, 2015). Stimulants are also known to enhance alertness, and some college students report that they use the medication to "get high" at parties (Benson et al., 2015). In fact, a recent meta-analysis of 20 studies found that rates of illegal use of stimulants by college students is estimated to be 17% (Benson et al., 2015). Another potential benefit of access to stimulants is the ability to sell the medication, and almost one-third of college-aged students reported diverting their prescribed stimulants illegally (Benson et al., 2015).

In addition to stimulants, college students hold positive attitudes about academic and testing accommodations that can be provided to individuals with ADHD. In one college sample ($n$ = 475), 67% of non-ADHD students reported that all students should have access to test accommodations, and many viewed extended time, private room, and extra breaks as benefits for high stakes testing situations (Lewandowski et al., 2014). The benefits of an ADHD diagnosis, coupled with the positive perceptions of such benefits, create incentivizing conditions that increase motivation to obtain the diagnosis (Musso & Gouvier, 2014).

The misallocation of therapeutic resources to malingerers is associated with increased societal costs and possible negative consequences. Colleges incur costs associated with specialized academic programing and test accommodations (Jasinski & Ranseen, 2011). Although exact rates are unreported, one can estimate that if ADHD accounts for one of the largest disability groups on college campuses (Raue & Lewis, 2011) and malingering is occurring at high rates among college ADHD evaluations (~25%; Marshall et al., 2010) then most likely some proportion of disability service costs are erroneously budgeted to those that truly do not need them. Furthermore, there are increased expenditures incurred by health insurance companies who cover medication prescriptions that are falsely obtained. It is estimated that insurance companies pay approximately $6.9 - 17 million dollars every 30 days on stimulant prescriptions that are eventually diverted (Aldridge, Kroutil, Cowell, Reeves, & Van Brunt, 2011). Beyond dollars spent, the illegal and unmonitored use of stimulants increases the potential for adverse health outcomes, such as increased risk of stimulant abuse/dependence and adverse medication side effects (Park & Haning, 2016). With increasing evidence and concern that feigned ADHD occurs in a significant proportion of college campus evaluations, a better understanding of the malingering response style could help to inform detection methods.

## Ability to Malinger ADHD

Research indicates that the ADHD test profile can be easily faked by individuals motivated to obtain the diagnosis. An ADHD diagnosis in adults often relies upon symptom reports, interviews, observations, and clinical judgment (Nelson et al., 2014). Studies have demonstrated that individuals can accurately feign ADHD symptoms, impairment, and executive dysfunctions on rating scales (Marshall et al., 2010, 2016; Quinn, 2003; Tucha et al., 2015). In addition, college students motivated to obtain the diagnosis can fake the ADHD cognitive profile, including measures of intelligence, verbal working memory, attention, and reading

(Booksh et al., 2010; Harrison, Rosenblum, & Currie, 2010; Marshall et al., 2010; Sullivan et al., 2007). Thus, a comprehensive ADHD assessment that includes rating scales and performance measures cannot effectively discriminate feigned from actual ADHD cases (see Tucha et al., 2015 for a detailed review).

Research has noted that feigned cases tend to exhibit an exaggerated response bias, in that they under-perform on cognitive tests and over-report on rating scales in comparison to individuals with ADHD (Harrison & Edwards, 2010; Marshall et al., 2010, 2016; Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008). Despite these response tendencies, there is rarely an obvious difference between scores of those with ADHD and those faking. Consequently, the profile of a feigner could look like a genuine ADHD profile (Edmundson et al., 2017; Fuermaier et al., 2016). The overlap between feigned ADHD and real ADHD test profiles makes it a challenge to differentiate these groups (Musso & Gouvier, 2014; Tucha et al., 2015).

Research also has found that one can fake the diagnosis with little to no preparation (Edmundson et al., 2017; Fuermaier et al., 2016). ADHD is a common neurodevelopmental disorder and the symptoms are known by the general population (McLeod, 2007). Such prior knowledge appears to be sufficient for an individual to fake the disorder on measures used to diagnose ADHD (Fuermaier et al., 2016; Tucha, Sontag, Walitza, & Lange, 2009). But, a true malingerer may come to an ADHD evaluation well-prepared on the diagnostic criteria, disorder profile, and even the psychological measures they might complete (Rogers, 2008). Research has found that even a brief review of ADHD information tends to improve an individual's ability to fake an ADHD profile that is often indistinguishable from true cases (Edmundson et al., 2017; Fuermaier et al., 2016). Information about ADHD can easily be obtained online. Such information includes the *DSM-5* diagnostic criteria, diagnostic measures (e.g., rating scales), and

even advice on how to fake ADHD. The ease that non-ADHD students can fake the disorder, combined with the ease to which one can access information about the disorder makes it a challenge to detect feigned ADHD (Musso & Gouvier, 2014; Rogers, 2008)

Assessment practices that emphasize self-report rating scales in the assessment of ADHD enhance the ease of faking the disorder (Harrison, 2017; Jasinski & Ranseen, 2011; Musso & Gouvier, 2014). Most ADHD rating scales have high face validity, and most scales only include items that simply replicate the *DSM-5* symptoms (Musso & Gouvier, 2014). Thus, even a brief review of the diagnostic criteria makes it easy for individuals to identify the specific items they need to endorse. Furthermore, most rating scales were designed to be screening measures with lower thresholds to rule-in as many probable cases as possible. These low thresholds make it easy for someone who is intentionally over-reporting symptoms to exceed the clinical cut score and be screened-in for diagnosis (Marshall et al., 2016). For example, Marshall and colleagues (2016) found that a significantly higher proportion of those suspected of malingering exceeded clinical cut scores on ADHD symptom, impairment, and executive functioning scales compared to clinical cases. In other words, rating scales are quite good at screening-in elevated symptoms, and even better at screening-in exaggerated reports.

Another complication with accurate assessment is that some individuals with ADHD tend to under-report symptom and impairment levels as compared to collateral reports (Dvorsky, Langberg, Molitor, & Bourchtein, 2016; Prevatt et al., 2012). The tendency for some with the clinical disorder to under-report concerns has been partly attributed to the positive illusory bias, which is the belief that one is more competent than actual skills and abilities indicate. While positive illusory bias may protect the individual from experiencing feelings of inadequacy (Owens, Goldfine, Evangelista, Hoza, & Kaiser, 2007), a lack of self-awareness could lead someone to minimize symptom and impairment levels on self-report rating scales (Owens et al.,

2007; Prevatt et al., 2012). In summary, both individuals with ADHD and those faking the disorder could report inaccurate information on rating scales. Thus, without assessing for validity of such self-reports, reliance on subjective rating scales alone could lead to invalid diagnoses.

Finally, malingering could go undetected due to clinician and diagnostic bias. Although the *DSM*-5 provides clear criteria to confer a disorder (≥ 5 symptoms), most clinicians report using flexible assessment practices to accommodate situations in which they "believe" the individual has the disorder, even though they may not have sufficient evidence of symptoms and impairment to support the claims (Harrison, 2017; Nelson et al., 2014). Such beliefs could reflect a confirmatory bias, or the desire to find information that supports initial clinical judgment, even if the data reflect poor effort or exaggeration. Furthermore, clinicians may want to provide individuals with a diagnostic explanation for the self-reported concerns, and overlook signs that the individual is malingering (McLaughlin, 2002). Flexible thresholds and diagnostic biases, combined with the easy ability to fake ADHD symptoms, increase the risk of false positive diagnoses. If traditional ADHD measures (e.g., rating scales) cannot easily detect feigned ADHD, there appears to be a need to identify specialized tests and methods that can detect this population.

## Malingering Detection

Validity tests are specialized measures designed to assess performance effort and response credibility in psychological evaluations (Rogers, 2008). Performance validity tests (PVT) assess for testing behaviors or effort on skill-based tasks (Larrabee, 2012). PVTs often use methods of extreme number of failed items (magnitude of errors) or disproportionate failure of easy items in comparison to difficult ones (performance curve; Larrabee, 2012; Rogers, 2008). On the other hand, symptom validity tests (SVT) assess for credibility in subjective reports. SVTs are often embedded into rating scales and use detection strategies of exaggerated symptom

reports (e.g., frequency, intensity), along with over-endorsement of infrequent symptoms, stereotypes, or rare symptom combinations (Larrabee, 2012; Rogers, 2008). The best performance or symptom validity tests often include specifically created malingering items or tasks with the specific intention of malingering detection. While these tests can be standalone/free-standing or embedded into current rating scales, it is important to clarify that these tests are different than embedded validity indicators. Validity indicators are created using existing items within diagnostic or ability measures (e.g., cognitive tests, continuous performance tasks). Oftentimes, these validity indicators include cut scores that reflect extreme response bias or poor performance on ability tests. For example, an extremely high score on a clinical symptom scale could reflect symptom exaggeration. Yet, because these validity indicators use items intentionally designed to detect clinical symptoms/impairment, and not items intentionally designed for malingering, these indicators are at higher risk of misclassifying an honest reporter with significant deficits/impairments as "feigning" (Heilbronner et al., 2009; Rogers, 2008).

Malingering is difficult to verify, at least in a definitive sense, in the absence of a confession. Thus, malingering classifications can only be suspected based upon the available evidence that supports the presence of intentional faking to obtain external benefits (Heilbronner et al., 2009; Slick, Sherman, & Iverson, 1999). Validity tests help to provide objective data that reflects a certain probability of malingering. These probabilities are derived from research studies on the validity test's ability to correctly classify malingering in the intended populations (i.e., detect malingered brain injury within the cognitive impairment population). Although different probabilities can be calculated, two commonly referenced probabilities are sensitivity and specificity. As it relates to malingering, sensitivity is the proportion of malingerers who were correctly identified based on a positive validity test (e.g., noncredible or fail). Specificity is the

proportion of honest reporters that were accurately ruled-out with a negative test outcome (e.g., credible or pass). Classification accuracy reflects the probability of correct test outcomes, and accounts for both sensitivity (true positives) and specificity (true negatives).

In order to reduce the chance that a validity test misclassifies an honest reporting individual as malingering, validity tests are designed to have high specificity rates (Larrabee, 2012; Rogers, 2008). Yet, the consequence of high specificity often is a reduction in sensitivity. For example, available validity tests designed to detect feigned brain injury demonstrate an average sensitivity rate of 69% at high specificity (90%; Sollman & Berry, 2011). Though validity tests are not perfect malingering measures, validity tests are generally more effective than clinical judgment or diagnostic measures alone (Booksh, Pella, Singh, & Gouvier, 2010; Rogers, 2008) For these reasons, validity tests are strongly recommended as part of evaluation contexts that involve the receipt of external incentives (Heilbronner et al., 2009; Iverson, 2006).

## Detection of Feigned ADHD

Although some ADHD rating scales include symptom validity tests or other embedded validity indicators (e.g., over-report cut scores), there are no published validity tests designed specifically to detect feigned ADHD. As such, the research community has attempted to repurpose existing measures to detect feigned ADHD. These attempts have included investigation of symptom and performance validity tests that are commercially available. Initially, researchers investigated those SVTs included in ADHD rating scales, even though those tests were not designed intentionally to detect malingering. Furthermore, studies have examined the use of PVTs to detect feigned ADHD, even though these tests were designed for use with feigned cognitive impairment and psychopathology. Recently, research has started to create detection measures for feigned ADHD, which will be discussed in more detail below.

### Published ADHD Validity Tests

ADHD is a symptoms-based disorder often diagnosed with use of self-report measures and rating scales (Nelson et al., 2014). Thus, research on feigned ADHD has investigated a variety of symptom validity tests (SVTs) included within published rating scales (Tucha et al., 2015), with the majority of these measures designed to detect feigned psychiatric conditions, such as feigned psychosis (Rogers, 2008). Unfortunately, these measures include items with little relation or face validity to the ADHD diagnosis (e.g., psychotic states, delusions, hallucinations); and consequently, these measures demonstrate little effectiveness to detect feigned ADHD (see Tucha et al., 2015). Although there are many ADHD rating scales, only two published measures include validity tests or indicators to assess for credibility of responding—The Conners' Adult ADHD Rating Scales (Conners, Erhardt, & Sparrow, 1998) and the Clinical Assessment of Attention Deficit-Adult (CAT-A; Bracken & Boatwright, 2005).

**The Conners' Adult ADHD Rating Scales.** The Conners' Adult ADHD Rating Scales (Conners et al., 1998) is a self-report measure that assesses ADHD symptoms, differential disorder symptoms (i.e., conduct, anxiety, depression), and includes three measures to assess response validity. These validity measures include the Inconsistency index, and two over-report indicators on the Inattentive scale and Hyperactive-Impulsive scale. The existing CAARS validity measures were designed to detect general respondent validity (e.g., inattention, negativity), not specifically feigned ADHD. As this self-report rating scale is used commonly in ADHD evaluations, a substantial amount of research has focused on whether these validity measures could detect this population.

***CAARS Inconsistency index*.** The CAARS Inconsistency index is a validity scale designed to detect consistency in responses between two similar symptom items, with the expectation that a valid responder would report similar responses on the same symptom questions (e.g., fidget in seat). Therefore, the Inconsistency index can be helpful to detect

inattention, lack of motivation, and/or misinterpretation of the item statement (Harp et al., 2011; Jasinski et al., 2011).

Suhr and colleagues (2008) conducted the first investigation on the ability of the CAARS Inconsistency index to detect feigned ADHD. The study examined archival data from college ADHD evaluations. The average age from the entire sample was 22.7 years (range 18–56 years). The researchers identified the ADHD group ($n$ = 15) as individuals who met the diagnostic criteria for ADHD and passed validity testing (the Word Memory Test). The Suspect group ($n$ = 26) included individuals who failed at least one validity test, presuming that this validity outcome reflected a feigned case of ADHD. The Clinical Control group ($n$ = 26) was comprised of individuals who passed validity testing, and who received a clinical diagnosis other than ADHD (e.g., depression). Results found that the Inconsistency index was insensitive to malingering, as all but three suspected malingerers responded consistently (Suhr et al., 2008). This finding that most malingerers are consistent responders has been corroborated in several archival and simulation studies (Edmundson et al., 2017; Fuermaier et al., 2016; Harrison & Armstrong, 2016; Hirsch & Christiansen, 2015), and suggests that the Inconsistency index has little clinical usefulness for malingering detection.

***CAARS over-report indicators.*** In addition to the Inconsistency Index, the CAARS manual recommends that a T-score > 80 on either Inattentive or Hyperactive-Impulsive scales could indicate excessive over-report that may reflect invalid responding. Because those malingering ADHD tend to exaggerate responding in comparison to ADHD cases (Marshall et al., 2016; Quinn, 2003), Suhr and colleagues (2008) also investigated whether the over-report indicators on the CAARS symptom scales could be used to detect cases of malingering. Utilizing the same data set that investigated the Inconsistency index, Suhr et al. (2008) found that the Suspect group did exhibit an exaggerated response bias, with suspected malingerers obtaining

higher scores on both Inattentive and Hyperactive-Impulsive scales in comparison to the ADHD and Clinical Control groups. Yet, despite this exaggerated response bias, the two over-report indicators were ineffective discriminators of suspected malingerers. Specifically, only 33% of Suspect malingerers exceeded the over-report cut score (T-score > 80) on the Hyperactive-Impulsive scale. On the Inattentive scale, there was no significant response differences between the Suspect and ADHD groups, with over half of participants within both groups exceeding a T-score > 80. These findings have been replicated in other investigations (Edmundson et al., 2017; Fuermaier et al., 2016; Harrison & Armstrong, 2016; Hirsch & Christiansen, 2015), and indicate that an elevated T-score on the CAARS symptom scales is an ineffective method of discriminating feigned cases from the clinical disorder.

***Conners' Infrequency Index*.** Research indicates that the existing CAARS Inconsistency index and over-report indicators should not be relied upon to detect feigned ADHD (e.g., Suhr et al., 2008). Nonetheless, it should be possible to formulate embedded validity indicators from existing items within rating scales and tests (Rogers, 2008). Given that the CAARS is a comprehensive rating measure that includes multiple different symptom items (e.g., learning, memory), one research team investigated whether an infrequency index could be formed from these existing CAARS items to detect cases of malingering (Suhr, Buelow, & Riddle, 2011).

In a two-part study intended to create a Conner's Infrequency Index (CII), the researchers used archival data of psychological evaluations of 1,173 individuals with an average age of 19 years (age range of 18-25 years) to form ADHD ($n$ = 77), Clinical Control (e.g., depression; $n$ = 147), and no diagnosis Control groups ($n$ = 955). Across all three groups, the researchers identified 12 CAARS symptom items that were endorsed infrequently (< 10%) in this first sample. After the 12-item CII was created, the researchers validated the ability of this index to detect malingering in a different archival data set comprised of honest reporting ADHD ($n$ = 19),

Suspect (*n* = 29), Clinical Control (i.e., learning disability, depression; *n* = 43), and Control group (i.e., no disability; *n* = 33). A performance validity test was used to classify the groups as Suspect (positive test), or honest reporter (negative test). The study findings indicated that the created CII could identify participants who over-reported (T-score >80) on the CAARS over-report indicators (Inattentive or Hyperactive-Impulsive scales); yet, the CII was only able to detect 24% of those who also displayed noncredible performance on the performance validity test (Suhr et al., 2011).

In a recent validation study for the CII, Fuermaier and colleagues (2016) analyzed the ability of this measure to detect non-ADHD adults instructed to simulate malingered ADHD. For this simulation study, non-ADHD adults between the ages of 18 - 58 years (*M* =27.5, *SD* = 11.0 years) were recruited from the community, and randomly assigned to different simulation conditions. The Naïve Simulation group (*n* = 87) received no preparation beyond a simple prompt to feign ADHD. In contrast, the well-prepared Symptom-Test Coached group (*n* = 91) was provided with five minutes to review information on ADHD symptoms and typical assessment process (e.g., tests involved in ADHD evaluations). The ADHD group (*n* = 52) consisted of adults who met the *DSM-IV* diagnostic criteria based upon agreement by two different psychologists.

Results from the validation study indicated that the CII could not effectively detect either the Naïve Simulation or Symptom-Test Coached groups (Fuermaier et al., 2016). For example, although the CII demonstrated moderate sensitivity (52%) for the unprepared, Naïve Simulation group, specificity was only 65% of the clinical ADHD group. The CII demonstrated even lower sensitivity (32%) to detect the well-prepared Symptom-Test Coached group (Fuermaier et al., 2016). Two additional studies also have found that that the CII has weak abilities to detect

feigned ADHD (Cook, Bolinger, & Suhr, 2016; Edmundson et al., 2017; Hirsch & Christiansen, 2015), and therefore, this created validity indicator is not recommended for clinical use.

***Conclusions on the CAARS validity tests.*** The studies reviewed above indicate that the existing CAARS validity indicators and the created CII have limited ability to detect feigned ADHD; however, they do demonstrate a relationship between feigned ADHD and an exaggerated response style. Specifically, there is a tendency for the malingering group to over-report symptoms, even on items that are endorsed infrequently by the clinical ADHD group (Suhr et al., 2011). Yet despite this symptom over-report, the use of ADHD symptom and even ADHD-related items are unable to reliably discriminate feigned ADHD (Cook et al., 2016; Fuermaier et al., 2016). As previously discussed, ADHD is a heterogeneous population that reports different symptoms at different degrees of severity (e.g., Barkley et al., 2008). Therefore, we expect individuals with ADHD to endorse ADHD symptom items, and sometimes ADHD-related ones. As a result, simple over-report on items used to diagnose ADHD is not recommended as a method to detect malingering (Musso & Gouvier, 2014).

**The Clinical Assessment of Attention Deficit-Adult: Infrequency scale.** The Clinical Assessment of Attention Deficit-Adult (CAT-A; Bracken & Boatwright, 2005) is another ADHD rating measure with a validity scale intended to detect response bias. The CAT-A Infrequency scale consists of 10 ADHD symptom items (e.g., "My friends cannot physically keep up with me") that were discovered in initial validation studies to be endorsed infrequently by the clinical population (< 6%). Therefore, the authors suggested that elevated reports (i.e., *strongly agree* or the highest response scale value) on these items could reflect a negative response bias (Bracken & Boatwright, 2005). Although this scale seems to have potential clinical utility, the authors did not confirm whether malingerers would actually respond highly on these infrequently endorsed

items. Therefore, the classification accuracy of the CAT-A Infrequency scale to detect feigned ADHD from clinical ADHD is best estimated from two research studies.

Marshall and colleagues (2010) conducted the first study on the CAT-A Infrequency scale's effectiveness to detect a group of suspected cases of feigned ADHD. The authors utilized a large archival data set of ADHD evaluations from a community-based neuropsychology practice ($n$ = 268) to analyze various different validity tests to detect suspected malingering, including the CAT-A. This study sample included a wide age range (17-59 years), but almost three-quarters of the sample were between the ages of 17-30. From the entire study sample, they identified suspected malingerers based upon positive outcomes on at least two validity tests. The ADHD group consisted of participants who met the *DSM-IV* criteria for ADHD, had cognitive testing consistent with the diagnosis, displayed impaired sustained attention on at least one measure, and passed the administered validity tests. The authors reported total sample sizes for each measure, but they did not report a specific sample size for each of these groups. From 167 cases of suspected malingerers and clinical cases of ADHD, they found that the CAT-A's Infrequency scale's manual cut score (≥ 3 items) had 58% sensitivity to detect the Suspect groups while accurately screening out 89% of the honest reporting ADHD group (specificity). A cut score ≥ 4 decreased sensitivity rates to 36%, but improved specificity (97%) to rule-out clinical ADHD (Marshall et al., 2010).

Given the possible utility of this measure to detect feigned ADHD, Marshall and colleagues (2016) incorporated the CAT-A in a follow-up study. This study combined a portion of archival data from the aforementioned 2010 study (166 participants) and included data from 262 individuals who completed an ADHD evaluation between March 2010 and July 2014. Using a criterion of ≥ 2 validity tests, Marshall and colleagues (2016) identified an honest reporting ADHD group ($n$ = 102), and a Suspect group ($n$ = 115). Corresponding to the aforementioned

study (Marshall et al., 2010), the researchers reported that 57% of the Suspect group exceeded the CAT-A Infrequency scale's manual cut score (≥ 3 items). Although this information provides additional information on the sensitivity of the CAT-A Infrequency scale to detect cases of suspected feigning (57 - 58%), the authors did not report any additional classification accuracy analyses, such as specificity of a clinical ADHD group.

The findings from both archival studies conducted by Marshall and colleagues (2010, 2016) suggest the CAT-A Infrequency scale could detect cases of feigning better than the CAARS validity tests. However with archival designs, researchers often use validity tests to determine whether cases reflect "suspected malingering" or "honest." There is no assurance that a "suspected" case actually is a "true" case of malingering (Rogers, 2008). To date, no other research has been conducted on the CAT-A. Given the inherent limitations of archival research that can only suspect malingering (and cannot prove malingering), additional research appears needed to validate the utility of the CAT-A to detect feigned cases of ADHD.

**Non-ADHD Performance Validity Tests**

As previously mentioned, there is no available performance validity test (PVT) designed empirically to detect feigned ADHD. However, because commercially available PVTs demonstrate high detection accuracy for feigning within their intended populations (e.g., feigned psychiatric symptoms or cognitive impairments), researchers questioned whether these measures could also detect feigned ADHD as well. Although some tests have been applied to ADHD (Tucha et al., 2015), only two measures appear to have some capacity (sensitivity > 50%) to detect feigned ADHD—the Word Memory Test (Green, 2003) and the Victoria Symptom Validity Test (Slick, Hopp, Strauss, & Thompson, 1996). Of these two measures, the WMT is used frequently to classify suspected cases of feigning in archival data of ADHD evaluations, operating under the presumption that the WMT is the best of the available performance validity

tests, and perhaps one of the only available options (Harrison & Armstrong, 2016; Harrison & Edwards, 2010; Suhr, Buelow, et al., 2011; Suhr et al., 2008; Suhr, Sullivan, & Rodriguez, 2011; Sullivan et al., 2007).

**Word Memory Test.** The Word Memory Test (WMT; Green, 2003) is a forced-choice PVT designed specifically to detect feigned memory deficits in those with suspected cognitive impairment (e.g., traumatic brain injury). In the computerized version, individuals read a list of 20 pairs of words twice. Next, they are asked to select the word in the original list from two response options immediately (Immediate Recognition; IR). After a 20-minute delay, examinees are administered another forced-choice trial (Delayed Recognition; DR). As an additional WMT variable, Consistency (CNS) reflects discrepancies in responses between IR and DR subtests. Across the three main WMT measures (IR, DR, CNS), an accuracy score ≤ 82.5 % on these variables is suspected to be feigned performance (Green, 2003; Green et al., 2011). The WMT has been shown to be an effective validity test of feigned cognitive impairment, with a recent meta-analysis reporting an average correct classification accuracy rate of 79% to detect feigned and clinical cases of cognitive impairment (*SD* = 13.6%; Sollman & Berry, 2011).

Although the WMT is a common measure used in feigned ADHD research, the classification accuracy of this measure can only be estimated from three studies. The first of those studies happens to be the same study that investigated the CAT-A Infrequency scale. Marshall and colleagues (2010) utilized assessment data collected from a community based neuropsychological practice. From a large archival data set (*n* = 268), 20% of individuals completed the WMT (*n* = 53), with an unreported number in each group (ADHD and Suspect). The results from Marshall and colleagues (2010) indicated that the WMT IR subtest and the CNS calculation each demonstrated 63% sensitivity to detect the Suspect group while maintaining optimal specificity of the clinical group (> 90%; Marshall et al., 2010). On the other hand, the

DR subtest, had high specificity (90%) for the ADHD group, but demonstrated very low sensitivity for the Suspect group (18%; Marshall et al., 2010).

Results from this study suggest that perhaps two WMT measures could detect over half of individuals feigning ADHD. However, the archival study conducted by Marshall et al., (2010) may be limited by the potential use of the WMT for both group classification and group differentiation (ADHD vs. feigned ADHD). Because the WMT was listed as one of the multiple validity tests used for group classification of archival data, it is possible that some proportion of the Suspect group was classified a priori as "suspect" based upon a positive WMT outcome before performing the actual planned analyses with the same measure. Without the use of external validity tests, it is unknown whether the results simply reflect that the WMT initially classified 63% of the Suspect group, or whether the WMT *can* identify 63% of individuals suspected of malingering.

Two simulation studies have also investigated the WMT's ability to detect malingered ADHD (Booksh et al., 2010; Edmundson et al., 2017). In the first simulation study, Booksh and colleagues (2010) randomly assigned 110 undergraduate students without ADHD (average age 20.4 years) to Simulated malingering ($n$ = 54) and Control groups ($n$ = 56). The Simulated group received instructions to fake ADHD and to avoid detection by avoiding exaggerated response bias (i.e., avoid responding to all items with highest/lowest value). The ADHD group was derived from archival data of 56 participants who completed a full psychoeducational evaluation and met *DSM-IV* criteria for ADHD diagnosis.

Booksh and colleagues (2010) found that failure on at least one WMT measure (≤ 82.5% on IR, DR, or CNS) had the ability to correctly classify 58% of the Simulators. However, the researchers did not specify the classification accuracy for the tests, thus, it is unknown whether they found all three measures to be effective, or only the WMT IR subtest and CNS calculation

as found by Marshall et al. (2010). Furthermore, the ADHD group ($n$ = 56) in this simulation study did not complete the WMT, thus the findings only reflect the proportion of malingerers identified as credible/noncredible based upon the cut score, not whether the WMT could differentiate clinical ADHD versus feigned cases. One could assume that clinical ADHD would pass a validity test that is passed by most. But, the WMT does require some degree of attention (i.e., focus on the presented word list), and it is possible that the WMT may be susceptible to mistakenly classifying a clinical case of ADHD as feigning. Thus, without the inclusion of a clinical ADHD group, the findings from Booksh and colleagues (2010) do not provide substantial evidence for the ability of the WMT to discriminate malingering from clinical ADHD.

More recently, Edmondson and colleagues (2017) investigated the WMT's ability to detect a group of individuals instructed to simulate malingered ADHD. The researchers also investigated whether coaching, or brief review of ADHD symptom information, could help a non-ADHD participant's ability to fake the diagnosis without detection as feigning (i.e., pass validity testing). In this study, the researchers randomly assigned a group of non-ADHD participants from an undergraduate psychology course to Non-Clinical Honest ($n$ = 9), Non-Coached Malingering ($n$ =23), and Coached Malingering ($n$ =23). A small ADHD group ($n$ = 21) consisted of individuals who self-reported they met the *DSM-5* criteria for ADHD. The average ages within the four groups were similar (average age of 18-19 years old). All individuals were administered the Adult ADHD Rating Scale (ASRS; Kessler et al., 2005) as a pretest before administration of group instructions. Results from the ASRS found that the ADHD group reported significantly more symptoms in comparison to non-ADHD participants, which the researchers used as evidence to support correct group assignment of ADHD and non-ADHD participants. Subsequently, the ADHD and Non-Clinical Honest groups received instructions to

respond honestly on the study's primary measures. Both Non-Coached and Coached Malingering groups were instructed to fake the ADHD diagnosis without being detected as a malingerer (i.e., avoid extreme response bias). But in addition, the Coached Malingering group was provided with the ADHD symptom criteria to review before completing the study.

The findings from this study found that the WMT had high specificity to rule-out clinical ADHD (86% - 95%). Yet, the WMT was less sensitive at detecting Coached cases of malingering. Specifically, though the IR subtest and CNS calculation were relatively effective to detect Non-Coached Malingering participants, with sensitivity of 70% and 74% respectively, those same measures were less able to classify Coached ones correctly (IR sensitivity = 43%; CNS sensitivity= 52%). Furthermore, consistent with results from Marshall et al., (2010), the DR subtest was the least effective WMT subtest, with moderate sensitivity for Uncoached participants (57%) and very low sensitivity for the Coached condition (30%). The findings from this study suggest that the WMT may be less effective at detecting individuals who are prepared to fake the disorder, even if such preparation is brief and focused on ADHD symptoms. Yet because this study is limited by a small sample sizes, the researchers of this study encouraged the need to validate the WMT ability in additional samples.

**Conclusion on the WMT.** The WMT is a commonly used validity test in neuropsychological and ADHD evaluations, presumably because there are few options available and there is some research to support its use (Booksh et al., 2010; Edmundson et al., 2017; Marshall et al., 2010). An archival study suggested that the WMT could demonstrate moderate sensitivity (63%) to detect a group of individuals classified with suspected poor effort during a psychological evaluation. Two simulation studies indicated that the WMT may be fairly sensitive to detecting uncoached, or unprepared malingerers (sensitivity = 57 - 74%). However, the WMT appears to be less effective to detect individuals instructed to fake ADHD in a manner to avoid

detection (58%; Booksh et al., 2010), and even less effective to detect individuals coached on the symptom criteria of the disorder (30%; Edmundson et al., 2017). Perhaps, coached malingerers know that poor memory is not an essential ADHD symptom, and they have no reason to fail the WMT, which is an obvious test of memory. This could explain the above findings and help explain why the WMT may not accurately discriminate malingering. The different findings between coached and uncoached malingerers suggests that additional research is needed to examine the utility of the WMT as a measure of feigned ADHD.

### Validity Test Designed for Malingered ADHD

Based on research suggesting that available validity measures are ineffective at detecting feigned ADHD, some research groups (e.g., (Fuermaier et al., 2017; Harrison & Armstrong, 2016; Potts, 2016) have set out to create a measure designed empirically to detect this group. These investigations have included formulating an Exaggeration Index comprised of items from various scales (Harrison & Armstrong, 2016), and creating a visual-spatial validity test to detect malingered ADHD (Fuermaier et al., 2017). In addition, a Symptom Validity Index embedded within a Multidimensional ADHD Rating Scale showed promise as a malingering measure (Potts, 2016). Given that these unpublished measures would benefit from additional research, this section will focus on the measure created by the present author and colleagues.

**Multidimensional ADHD Rating Scale.** Research indicates that diagnostic accuracy of ADHD should improve with the assessment of symptoms, impairment, and malingering (e.g., Gathje et al., 2008; Harrison, 2017; Tucha et al., 2015). The assessment of symptoms and impairment helps to increase detection of clinical cases of ADHD from non-ADHD cases; and the assessment of respondent validity helps to rule-out cases of feigning. With this in mind, the present author and associates set out to create a comprehensive rating scale that targets these three areas. The Multidimensional ADHD Rating Scale (MARS) includes 18 ADHD *DSM-5*

symptom items and 22 functional impairment items to assess two criteria needed for clinical diagnosis, as well as 104 ADHD-like symptom validity items that were suspected to be frequently endorsed by malingerers, but not endorsed by those with the clinical disorder.

The MARS has been investigated in one pilot study (Potts, 2016). This simulation study randomly assigned a group of non-ADHD undergraduate students to Control ($n$ = 62) and Malingering ($n$ = 56) groups. The Malingering group was provided a malingering scenario used in prior research (Sollman, Ranseen, & Berry, 2010), and 5-minutes study time with general ADHD symptom information. The ADHD group ($n$ = 39) was recruited from the community and underwent a phone screening to corroborate a professional diagnosis.

In an analysis of the 104 experimental symptom validity items, seven items were found to have good discriminative ability to detect the simulated Malingering group from clinical ADHD. Using a cut score that favored optimal specificity (~90%), the 7-item Symptom Validity Index (SV-index) demonstrated better sensitivity (79%) than any other available validity test (e.g., WMT, CAT-A). Results suggested that a validity test created to have high face validity for a specific disorder, with items infrequently endorsed by an ADHD group, seemed to accurately detect feigned cases in college students. A necessary next step would be to see if this finding could be replicated.

As a measure of ADHD, the MARS was able to differentiate ADHD from non-ADHD controls with high accuracy on the MARS symptom indices (accuracy = 77 – 80%) when using a higher cut score. However, specificity was very low (33-53%) on symptom and impairment indices with the use of a lower threshold that favored sensitivity for ADHD, presumed to be because the researchers included a mixed control group comprised of individuals with comorbidities and ADHD characteristics. The use of a verified non-ADHD group and ADHD group may better validate the effectiveness of the MARS indices to detect ADHD from controls.

While the primary focus of the MARS clinical indices is to detect ADHD, further analysis of the MARS indicated that malingered ADHD may be detected, to a degree, via over-report on the ADHD symptom indices. In Potts (2016), the Malingering group had more extreme scores on the MARS Hyperactivity-Impulsivity index (HI-index), such that the use of an "over-report" cut score ≥ 49.5 had high sensitivity (75%) to detect malingering at optimal specificity of ADHD (~90%). Additionally, an over-report cut score of ≥ 55.5 on the Inattention Index (I-index) was able to detect 63% of Malingering participants at optimal specificity for clinical cases. Although Potts (2016) found that ADHD symptom indices could detect the exaggerated response bias associated with malingering, this finding is contrary to some prior studies (e.g., Suhr et al., 2008). One possible explanation is that the MARS has a 9-point response scale, while other ADHD rating scales (e.g., CAARS) use a shorter, 4-point response scale. As shorter response scales can experience a restriction of range effect (Preston & Colman, 2000), it is possible that the expanded response scale was better able to differentiate the slight over-report demonstrated by those feigning the condition. However, without additional research, it is unknown whether such an explanation is accurate, and whether the expanded MARS response scale improves detection accuracy of feigned ADHD.

Findings from this preliminary study suggested the MARS could be a useful tool to assess for ADHD symptoms, impairment, and symptom validity simultaneously. Results also supported that improved detection accuracy of malingered ADHD could occur with symptom validity items that are specifically created for this purpose. Furthermore, an expanded response scale could be useful to detect cases of noncredible over-report on ADHD clinical indices. Although promising, this study is limited by the use of the same sample to identify the SV-index as well as to conduct classification accuracy analyses. As such, the reported findings of the SV-

index are most likely over-inflated, and this measure is in need of replication on different samples.

**Aggregation of Multiple Validity Tests**

Malingering determinations should not be made lightly. Such decisions have clinical and legal consequences. For example, an individual classified as malingering can be denied treatment and benefits designed to improve the quality of life for those with clinical impairments. Because there is a certain degree of probability that an honest reporter achieved a positive outcome on a validity test (~10%), the field of neuropsychology recommends the use of multiple validity tests to increase accuracy of correct malingering classifications (Bush et al., 2005; Heilbronner et al., 2009; Larrabee, 2012).

Studies on feigned ADHD have demonstrated the increased probability of correct classifications with use of multiple test outcomes. In illustration, Edmondson and colleagues (2017) aggregated nine validity measures derived from the WMT, the CAARS, along with two other neurocognitive validity tests. Each positive test outcome contributed one point towards a total "aggregated" score, and classification accuracy calculations were calculated at each cut score (≥ 1 positive test; ≥ 2 positive tests). With this aggregation method, they found that multiple positive outcomes across these measures increased specificity to rule-out honest reporting cases. In fact, specificity went from 62% with ≥ 1 positive tests to 90% with ≥ 3 positive tests. But, sensitivity to detect malingering decreased from 65% with ≥ 1 positive tests, to 39% with ≥ 3 positive tests (Edmundson et al., 2017). However, classification accuracy of validity test combinations most likely is contingent upon the quality of each included measure. As such, it is possible that a different combination of validity tests could improve classification of malingering.

Harrison and Armstrong (2016) conducted a study that combined multiple embedded validity indicators from two diagnostic measures to create a weighted validity index. They utilized eight indicators from the two measures: Five dissociative symptom items (1 point each = 5 points), the total score on 18 dissociative symptom items (1 point), along with a T-score ≥ 80 on the CAARS Inattentive and/or Hyperactive-Impulsive scales (1 point each = 2 points). Each positive validity indicator received one point, and classification accuracy rates were conducted at each cut point (≥ 1 through 8 positive indicators). Results found that the weighted validity index yielded increased probability of correct classifications compared to the standalone ability of each validity indicator alone. In this appropriately named Exaggeration Index, a cut score of ≥ 1 positive validity indicator had moderate sensitivity (51%) to detect a combined Suspect/Simulation group at optimal specificity (88%) to rule-out ADHD. A cut score of ≥ 2 positive validity indicators had reduced sensitivity (34%), but increased specificity of ADHD (94%; Harrison & Armstrong, 2016). Although adopting a higher threshold demonstrated weaker sensitivity, the ability to rule-out honest reporters with high accuracy (specificity) ultimately equates to increased clinical confidence that this outcome reflects malingering.

## Purpose of the Study

Evidence suggests that a substantial minority of adults who self-refer for an ADHD evaluation are malingering, with estimates ranging from 20% to 50% (Harrison & Edwards, 2010; Marshall et al., 2010; Sullivan et al., 2007). With little preparation, motivated individuals can easily feign an ADHD diagnosis on psychological measures and on rating scales (Marshall et al., 2016). To date, the available measures that demonstrate some effectiveness (i.e., sensitivity and specificity) do not have sufficient research evidence to support their usage with feigned ADHD. The current study examined the ability of a newly constructed Multidimensional ADHD Rating Scales (MARS) to accurately differentiate true cases of ADHD from malingered ADHD

and non-ADHD controls. In particular, the composition and effectiveness of the MARS Symptom Validity Index and MARS over-report indicators were evaluated independently and compared to the Word Memory Test and the CAT-A Infrequency scale. The MARS, WMT, and CAT-A were examined in combination to see if they could improve detectability over any of the measures alone. The MARS clinical indices were also evaluated to determine their ability to detect ADHD from honest reporting non-ADHD controls. In addition to these general purposes, the following research aims are detailed.

**Aim 1: Detection of Malingering**

**Aim 1A: Validate the original SV-index.** Due to the exploratory nature in which the SV-index was created (Potts, 2016), the first aim of this study was to replicate the utility of the seven items included within the SV-index. It was expected that the 7-item SV-index would emerge as an accurate discriminator of malingered ADHD in comparison to the other included validity measures, and would have correct classification accuracy > 80%.

**Aim 1B: Create and validate a revised MARS SV-index.** An additional aim of this study was to determine if a revised SV-index could improve detection over the original version. An exploratory analysis of all SV items was conducted to find any additional items that could be used to detect feigned ADHD. Any revised version of the SV-index would be expected to have better classification accuracy compared to the original SV-index.

**Aim 1C: Validate the MARS over-report indicators.** While the primary focus of the MARS clinical indices is to detect ADHD, higher cut scores on these indices could be used as potential indicators of credible/noncredible self-report. Thus, the next goal was to replicate the over-report indicators on the MARS clinical indices (i.e., symptoms and functional impairment) to detect the exaggerated response bias of malingering. Based upon prior research (Potts, 2016), it was expected that the use of MARS over-report indicators (i.e., higher cut scores) on the

MARS HI-index, I-index, and Total Symptom index would continue to demonstrate high effectiveness (classification accuracy > 70%) to discriminate cases of Malingering from diagnosed ADHD.

**Aim 1D: Comparison of validity tests to classify malingering.** The detection accuracy was compared among the MARS, WMT and CAT-A. Based on prior research, it was expected that the original or a revised SV-index to be the most accurate discriminator, followed by the MARS over-report indicators (HI-index, I-index, and FI-index), the WMT subtests, and the CAT-A Infrequency scale.

**Aim 1E: Combined use of multiple validity tests.** Based on the recommended use of multiple validity tests to classify malingering (Heilbronner et al., 2009; Musso & Gouvier, 2014), this aim examined the ability of multiple validity tests to discriminate cases of simulated malingering from ADHD cases. First, this study addressed whether the SV-index and the MARS over-report indicators could be combined into a MARS weighted validity index. Participants received one point for a positive/noncredible outcome on the SV-index and three MARS over-report indicators, and classification accuracy calculations were completed at each cut point (≥ 1 to 4 positive outcomes). It was expected that a MARS weighted validity index would have higher specificity to rule-out ADHD and higher correct classification accuracy than the individual ability of the SV-index and the MARS over-report indicators.

In addition, the detection of malingering was examined by aggregating nine validity tests (subtest, indicator, index) from the MARS, WMT, and CAT-A. Each positive/noncredible outcome on a validity test contributed one point toward an aggregated score (maximum = 9 positive test outcomes). Classification accuracy calculations were conducted at each successive cut point. It was expected that aggregating multiple validity tests would yield higher specificity to rule-out cases of clinical ADHD compared to one test alone.

**Aim 2: Detection of Clinical ADHD**

The development of a comprehensive rating scale for ADHD diagnosis presumes that the clinical indices will be effective in detecting true positives (i.e., those with ADHD). This aim investigated the ability of the MARS symptom indices and FI-index to differentiate cases of ADHD from honest reporting non-ADHD controls. It was expected that the MARS indices would each have high sensitivity and classification accuracy (> 70%) to discriminate ADHD from non-ADHD control participants. It was also expected that the combined use of symptom and impairment indices would increase correct classification accuracy for differentiating clinical ADHD from non-ADHD controls compared to the symptoms indices alone.

**Aim 3: Additional Validation of the MARS**

The final aim was to assess the internal consistency and convergent validity of the newly created MARS symptom, impairment, and symptom validity indices. Internal consistencies were estimated using Cronbach's alpha. Evidence of validity for the MARS was examined with correlations between MARS indices and the CAT-A and WMT.

## Method

**Participants**

Following approval from Syracuse University Institutional Review Board, this study recruited individuals with and without ADHD to form three groups: ADHD, Malingering, and Control groups. The following mathematical formula was used to estimate sample size ($n_{sp}$) in order to achieve a pre-determined specificity (Sp) for a single diagnostic test with a dichotomous outcome (Hajian-Tilaki, 2014):

$$n_{Sp} = \frac{Z_{\frac{\alpha}{2}}^2 \widehat{Sp}(1 - \widehat{Sp})}{d^2 \times (1 - Prev)}$$

In this formula, *Prev* reflects the prevalence of the clinical disorder. $\widehat{Sp}$ is the specificity value that the study seeks to obtain. $Z_{\frac{\alpha}{2}}$ reflects the z-score value of the confidence interval (1-α), and $d^2$ is the margin of error for the pre-determined value of specificity. As the formula applies to this study, $\widehat{Sp}$ equals the pre-determined optimal specificity value of 90% of adult ADHD, which has an estimated prevalence rate of 5% (Kessler et al., 2006; Weyandt & DuPaul, 2006). The confidence interval was set with an alpha level of .05, reflecting $Z_{\frac{\alpha}{2}}$ = 1.96. The margin of error ($d^2$ = .10) was selected to an achieved specificity of 90% ± 10%:

$$\text{Sample size } n_{sp}\,36 = \frac{(1.96)^2\,(.90)(1-.90)}{(.10)^2\,(1-.05)}$$

The formula suggested the anticipated ADHD group sample size was 36. Post-hoc analyses with G*Power and one-way analyses of variance confirmed the sample size of 36 had sufficient power (>.90) to detect effects across all primary measures.

**ADHD group.** The ADHD group was recruited via an undergraduate psychology course and publically posted flyers (Appendix B). In terms of eligibility criteria, participants were required (a) to have a professional diagnosis of ADHD, (b) to be between the ages of 18-26, and (c) to be English-speaking. Fifty-one undergraduate students registered for the study via SONA Subject Systems at Syracuse University. Four individuals initiated contact for more study information as a result of the study flyer.

Self-reported ADHD diagnoses were corroborated in two ways. First, individuals with self-reported diagnoses of ADHD completed an eligibility screening to verify diagnosis of ADHD. The four community participants completed the screening over the phone, and scheduled a time to complete the study after they were deemed eligible. The 51 individuals recruited via SONA Subject System completed the screening in-person during the assigned timeslot. Because

this was done in-person during the participant's assigned session. All SONA participants were allowed to participate to receive course credit, and data collected from participants with questionable diagnoses were later excluded from the data set ($n$ = 4). In addition to the eligibility screening, an external ADHD rating scale (CAT-A) was used to independently validate the presence of ADHD symptoms both currently and retrospectively from childhood. Confirmation of ADHD was defined as elevated symptom reports (T score ≥ 60; Mild Clinical risk) on the CAT-A Clinical Index, and at least one CAT-A Current Symptom scales (Inattention, Hyperactivity, or Impulsivity scales).

Of the 55 ADHD participants, 11 individuals were removed from analyses for reporting Normal range symptoms on all CAT-A Current Symptom scales and/or the CAT-A Clinical Index. Six additional ADHD participants were removed for the following reasons: negative eligibility screening ($n$ = 3), experimenter observed poor effort (i.e., rushed through materials; $n$ = 2), and extreme response bias (i.e., extreme response scale option on most items; $n$ = 1). A CONsolidated Standards Of Reporting Trials (CONSORT) diagram that reflects the process of study enrollment and data exclusion of ADHD participants can be found in Figure 1.

The final ADHD group ($n$ = 38) included 20 males (52.6%) and 18 females (47.4%) with an average age of 19.05 years ($SD$ = 1.86). Of the 27 ADHD participants who reported a college grade point average (GPA), the average was 2.65 ($SD$ = 1.24). All participants self-reported a primary diagnosis of ADHD. Sixteen (42.1%) reported at least one comorbid disorder, which included learning disability ($n$ = 10, 26.3%), anxiety and depression ($n$ = 5, 13.2%), and anxiety ($n$ = 1, 2.6%). While ADHD subtype was not reported or assessed in this study, the distribution of symptoms by type (i.e., inattentive, impulsive, and hyperactive) is presented for review. The CAT-A Current Symptom scales and the Clinical Index is located in Table 1.

**Non-ADHD groups.** Non-ADHD undergraduate students were recruited through the SONA Subject Pool, and randomly assigned to Malingering and Control groups. The eligibility criteria for Non-ADHD participants included: (a) no diagnosis of ADHD, (b) be between the ages of 18-26, and (c) English-speaking. The study initially recruited and 210 non-ADHD participants and randomly assigned them to Malingering and Control groups.

Of the 108 participants assigned to the Malingering group, 40 were removed from the analyses for the following reasons: Technical error prevented completion of WMT ($n$ = 5), missing data > 95% ($n$ = 3), self-disclosed poor adherence to the Malingering assignment (i.e., indicated they did not fake ADHD; $n$ = 16), failed ≥ 1 catch validity item embedded within the study survey ($n$ = 14), and extreme response bias ($n$ = 2). This study sought to have a Malingering group comprised of individuals who could successfully fake ADHD. Because an ADHD diagnosis requires both childhood onset and current symptoms, CAT-A data were used to verify that those instructed to simulate ADHD reported symptoms retrospectively and currently. Following the same method used with the ADHD group, 11 individuals were removed from analyses because they did not report elevated symptom levels (T score ≥ 60; Mild Clinical risk) on both the CAT-A Clinical Index and at least one CAT-A Current Symptom scale (Inattention, Hyperactivity, or Impulsivity).

Of the 102 participants randomly assigned to the Control condition, 32 were removed from the analyses for the following reasons: Technical error prevented completion of WMT ($n$ = 3), missing data> 95% ($n$ = 3), self-disclosed poor effort for the study ($n$ = 11), failed ≥ 1 catch validity item embedded within the study survey ($n$ = 8), and demonstrated extreme response bias ($n$ = 7). This study aimed to form a non-ADHD Control group. Although it is expected for non-ADHD individuals to report some ADHD symptoms, it is quite possible that individuals who report a high degree of symptoms are false negative cases (undiagnosed ADHD individuals), and

a decision was made to remove these cases to ensure a non-ADHD control group. Therefore, while participants who only reported elevated symptoms on one CAT-A Symptom scale were retained, 15 non-ADHD Controls were removed from study analyses for reporting elevated symptoms (T score ≥ 60; Mild Clinical risk) on the CAT-A Clinical Index, a total score derived from Current and Childhood scales, and one or more CAT-A Current Symptom scales (Inattention, Hyperactivity, or Impulsivity scales). A CONSORT diagram that reflects the study enrollment, randomization, and data exclusion of non-ADHD participants is located in Figure 2.

The final Malingering group ($n$ = 57) included 19 males (33.3%) and 38 females (66.7%) with an average age of 18.51 ($SD$ = 0.76). Of those that reported college grade point average ($n$ = 38), the average was 3.42 ($SD$ = 0.37). Most participants did not report a disability ($n$ = 49, 86%). Seven students disclosed a disability, including depression ($n$ = 4), anxiety ($n$ = 1), depression and anxiety ($n$ = 1), and hearing impairment ($n$ = 1). The final Control group ($n$ = 55) included 20 males (36.4%) and 35 females (63.6%) with an average age of 18.71 ($SD$ = 0.85). The average self-reported GPA for those that disclosed ($n$ = 34) was 3.43 ($SD$ = 0.33). The majority of non-ADHD controls did not disclose a disability ($n$ = 86; 90.9%). Of those individuals that disclosed a disability, four students reported the following disabilities: Learning disability ($n$ = 1), anxiety ($n$ =2), and medical condition ($n$ = 1). Additional group characteristics information is summarized in Table 2.

**Group characteristic analyses.** Chi-square tests were used to explore demographic characteristics between the groups. For all three groups, chi-square analyses were conducted on the demographic categories of sex, ethnicity, year in school, and reported first language (English/other), and one-way analyses of variance (ANOVA) were conducted on the demographic variables of age and GPA. Alpha was set at .05. There was no significant difference between the groups on sex, $\chi^2$ (2, $N$ = 150) = 3.88, $p$ = .14, ethnicity, $\chi^2$ (10, $N$ = 149) = 11.75, $p$

= .30, year in school, $\chi^2$ (8, $N$ = 149) = 8.91, $p$ = .35, and first language, $\chi^2$ (2, $N$ = 149) = 3.38, $p$ = .19. A one-way analysis of variance also revealed no significant differences between the groups on age, $F$ (2, 147) = 2.49, $p$ = .09. However, there were significant differences between the groups on GPA, $F$ (2, 96) = 11.66, $p$ < .001. Tukey post-hoc analyses revealed that the ADHD group reported a significant lower GPA ($M$ = 2.65, $SD$ = 1.24) than both non-ADHD groups. There were no significant differences in self-reported GPA between the Malingering group ($M$ = 3.42, $SD$ = 0.37) and Control group ($M$ = 3.42, $SD$ = 0.33). While GPA could be correlated with ADHD symptoms and impairment levels, self-reported GPA should not affect how an individual completes a rating scale or performs on validity tests. As such, GPA was not considered as a possible covariate in any analyses.

Additional chi-square analyses were conducted to determine whether there were differences between the two non-ADHD groups on sex, disability status, receipt of college academic accommodations, and current school problems. Results indicated no significant differences between non-ADHD groups on sex, $\chi^2$ (1, $n$ = 122) = .11, $p$ = .84, disability, $\chi^2$ (6, $n$ = 97) = 8.51, $p$ = .20, accommodations, $\chi^2$ (1, $n$ = 112) = .002, $p$ = .96, and school problems, $\chi^2$ (1, $n$ = 112) = 2.67, $p$ = .10.

**Materials**

**Multidimensional ADHD Rating Scale (MARS).** The MARS is a pilot measure designed to assess ADHD symptoms, impairment, and symptom validity (Potts, 2016). Section one contains the 18 ADHD symptoms (#1 - 18), 7-item SV-index (#19 - 25), and 76 experimental symptom validity items identified as plausible malingering items from prior research (Potts, 2016). Section one also contains three catch validity items that are embedded into the measure to assess for attention and study effort (e.g., “respond 3 if you are still reading this survey”). Section two contains functional impairment items (22 items). The MARS uses a 9-

point response scale, which research suggests could have higher discriminative power compared to shorter scales (Preston & Colman, 2000). Five response labels are equally spaced on a 0 - 8 numeric scale. Symptom and symptom validity items use a frequency scale (*Never*, *Rarely, Sometimes, Often,* and *Very Often),* while impairment items utilize a severity scale (*Not at All, Somewhat, Mild, Moderate,* and *Severe*).

The MARS is comprised of four clinical indices derived from symptom and impairment items to form an Inattention index (I-index), Hyperactivity-Impulsivity index (HI-Index), Total Symptom index, and Functional Impairment index (FI-index). Potts (2016) found that the following cut scores on these clinical indices had optimal sensitivity (~90%) to detect clinical ADHD from non-ADHD controls: I-index ≥ 20.0, HI-index ≥ 12.5, Total Symptom index ≥ 33.5, and FI-index ≥ 27.5. In addition, this research resulted in the creation of a Symptom Validity index. Furthermore, the study identified the possible use of over-report indicators (i.e., higher cut scores) on the symptom and impairment indices to detect cases of malingering. To reduce false positives decisions, this study utilized cut scores that favored optimal specificity (~90%) in the prior study (Potts, 2016): SV-index ≥ 28.0, I-index over-report ≥ 55.5, HI-index over-report ≥ 49.5, Total Symptom index over-report ≥ 100.5, and FI-index ≥ 127.5.

Only one study has been conducted on the MARS (Potts, 2016). In a prior study, the MARS indices demonstrated good internal consistency: Total Symptom index (α = .93), HI-index (α = .87), I-index (α = .92), and the SV-index (α = .84). But to date, no validation studies have been conducted on this measure.

**Clinical Assessment of Attention Deficit-Adult.** The Clinical Assessment of Attention Deficit-Adult (CAT-A; Bracken & Boatwright, 2005) is a 108-item adult ADHD self-report measure that includes a Childhood Memories section (54 items) and Current Symptoms section (54 items). Childhood Memories items and Current Symptom items form three separate

symptom scales: Inattentive (ATT), Hyperactivity (HYP), and Impulsivity (IMP). In addition, three Clinical Index scales are formed to provide summary scores for the Childhood Memories section, the Current Symptoms section, and a total score across both sections (The CAT-A Clinical Index). Raw scores are converted into T-scores, and the following classification labels are used to describe the relative risk of ADHD: Normal range (T score ≤ 59), Mild clinical risk (T score = 60-69), Significant clinical risk (T score = 70 - 79), and Very Significant clinical risk (T score ≥ 80) for ADHD. While normative data for the CAT-A begins at age 19, the CAT-A author (Bruce Bracken, PhD) indicated that the CAT-A clinical indices could be used for research on 18-year-old college students given that these young adults are most likely experiencing the same events as the young adults included in the CAT-A normative sample (e.g., emerging adulthood, independent/college life; B. Bracken, personal correspondence February 2017).

The CAT-A also assesses response validity with the Infrequency scale, which consists of 10 items (Childhood Memories items #10, 27, and 28; and Current Symptom items #58, 72, 77, 81, 83, 85, and 92) that were endorsed infrequently by both the clinical ADHD group (≤ 6%) and non-clinical general population (≤ 1%) during the standardization process. The manual indicates high endorsement (*strongly agree*) on ≥ 4 items may indicate noncredible responding. Marshall and colleagues (2010) reported the test manual's cut score of ≥ 4 had sensitivity of 22% for an archival sample of suspected malingering. The study identified that a lower cut score of ≥ 3 increased sensitivity to 58% for this population (Marshall et al., 2010).

The CAT-A manual reports validity and reliability information for the clinical indices. Specific to the scales used in this study, the Childhood Memories Clinical Index, Current Symptoms Clinical Index, and CAT-A Clinical Index demonstrated good internal consistency (α >.90). In a validation study, the authors reported the CAT-A scales had a correct classification

accuracy between 79 to 88% for the Symptom scales to differentiate clinical ADHD from learning disabilities and non-ADHD, non-disabled controls (Bracken & Boatwright, 2005).

Evidence of convergent validity was reported between the CAT-A and the Conners' Adult ADHD Rating Scales (CAARS; Conners, 1998). The authors reported the CAT-A Current symptom scales (Inattentive, Hyperactive, and Impulsivity) demonstrated moderate to high correlations ($r$ = .62 -.79) with the CAARS symptom scales (Inattentive and Hyperactive scales). Furthermore, strong correlations were found between the two scales' total symptoms indices—the CAT-A Current Symptoms Clinical Index and the CAARS ADHD Index ($r$ = .72) and the CAT-A Clinical Index and the CAARS ADHD Index and ($r$ = .69). Furthermore, evidence of convergent validity was reported for the CAT-A and the Brown Attention Deficit Scales, with the Brown Total Score correlated highly both the Current Symptoms Clinical Index ($r$ = .70) and the CAT-A Clinical Index ($r$ = .66). Lastly, divergent validity was reported between the CAT-A and the Clinical Assessment of Depression (CAD; Bracken & Howell, 2004), with correlations ranging from $r$ = .36 to .50 between the CAD scales (Depressed Mood, Anxiety/Worry, Diminished Interest, Cognitive and Physical Fatigue) and the CAT-A Current symptom scales, Current Symptom Clinical Index, and CAT-A Clinical Index (Bracken & Boatwright, 2005).

**Word Memory Test.** The Word Memory Test (WMT; Green, 2003), Windows version, is a performance validity test in which individuals are twice presented with 20 word pairs. Next, they select the word in the original list from 40 forced-choice Immediate Recognition (IR) subtest. After a 30-minute delay, they are administered a 40 forced-choice subtest of Delayed Recognition (DR). Consistency (CNS) is the calculation of reliability between responses on IR and DR subtests. Following the two primary subtests (IR and DR), the present study also administered the supplemental 6-choice Multiple Choice (MC) subtest, which is considered to be slightly more challenging compared to the forced-choice trials due to the possibility of semantic

interference with the six response options. As such, the MC subtest tends to have more variable specificity rates and lower overall classification accuracy to differentiate malingering from those with significant cognitive impairments (Green, 2003; Strauss, Sherman, & Spreen, 2006).

Validity and reliability information on the WMT has only been reported with feigned cognitive impairment. When used with this population, the WMT demonstrates high specificity (> 95%) for individuals with true cognitive impairment with a cut score of 82.5% correct (Green, 2003). The WMT subtests have high internal consistency (~90%) and high intercorrelations ($r$ = .80). Test retest reliability is modest at best ($r$ = .33 and .43) for IR and DR subtests, which is attributed to natural variability in individual effort levels, even in honest reporters (Strauss, Sherman, & Spreen, 2006). The WMT displays good convergent validity with other measures of effort, such as the Test of Memory Malingering ($r$ = .68; Strauss et al., 2006)

As a feigned ADHD detection measure, Marshall and colleagues (2010) found the manual cut scores ≤ 82.5% on the WMT IR or CNS each demonstrated 63% sensitivity, while the DR subtest had 18% sensitivity, to detect suspected cases at optimal specificity (> 90%). A simulation study found the WMT may be less effective to detect a malingering group coached on ADHD symptoms, with the reported sensitivity for WMT IR = 43%, WMT DR = 30%, and WMT CNS = 52%, and the specificity rates were 95%, 95%, and 86%, respectively (Edmundson et al., 2017). To date, no studies have evaluated the MC subtest's ability to differentiate feigned ADHD from clinical cases.

**ADHD screening form.** A screening form was used to verify study eligibility for individuals with ADHD. The focus of the screening questions was to verify ADHD diagnosis, along with current symptoms and impairment. Study eligibility was determined if they (a) have a diagnosis of ADHD, (b) received the diagnosis from a qualified professional (e.g., psychologist, counselor), (c) reported symptoms occurred before the age of 12, (d) experienced symptoms

currently, and (e) reported impairment in at least one area (i.e., academic, occupational, or social; Appendix C).

**Demographic questionnaire.** A demographic questionnaire was used to collect background information on participants. Questions included: (a) age, (b) gender, (c) ethnicity, (d) year in college, (e) GPA, (f) diagnosis, (g) school problems, (h) accommodations, (i) medication, (j) lifetime use of stimulant medication, (k) whether they believe they have ADHD, and (l) whether they have previously sought out an ADHD evaluation (Appendix D).

**Exit survey**. At the end of the study, individuals completed an exit questionnaire as a manipulation check to assess for study effort and adherence to the assigned group condition (honest or simulated malingering). Participants were also asked general questions about their prior knowledge of ADHD. The ADHD and Control groups' exit survey can be found in Appendix E. The Malingering group's exit survey contains additional items about strategies that they used during the self-report measure to successfully fake ADHD, and is located in Appendix F.

**Debriefing letter.** In order to increase effort, the groups were informed at the beginning of the study that they needed to put forth best effort (ADHD and Control) or successfully fake ADHD (Malingering) to receive an incentive ($100 Visa gift card raffle drawing). The debriefing letter informed participants at the end of the study that all individuals who completed the study, regardless of effort, would be entered into the raffle drawing.

## Procedures

**ADHD group.** Individuals with ADHD completed the study individually in a private testing room. Prior to participating in the study, individuals with ADHD completed the eligibility screening with the primary researcher or trained research assistant. Following the screening questions, the participant was seated at a computer to complete the online rating scales and

computerized WMT. The participant began the study by reviewing the electronic informed consent, and then completing the demographic questionnaire. ADHD participants were instructed to complete all the measures honestly and to respond to questions as if they were off ADHD medication. Next, the WMT IR subtest was administered, followed by the online rating scales in counterbalanced fashion: (a) MARS symptom and experimental symptom validity items, (b) MARS functional impairment items, (c) the CAT-A full measure. Twenty minutes after the completion of the IR subtest, participants were administered the WMT DR and MC subtests. After finishing the WMT subtests and rating scales, participants completed the final exit survey to conclude the study. The entire study, including the eligibility screening, took approximately 45-60 minutes.

**Non-ADHD groups.** The non-ADHD groups completed materials in small groups of 10 in a reserved computer lab. The non-ADHD study protocol followed the same procedures as outlined in the ADHD group, but using a blind procedure with two researchers. The first researcher administered the informed consent followed by the demographic survey. Subsequently, participants received an enclosed packet with information on group assignment and a set of instructions. Order of assignment was randomized prior to each session. The Control group received written instructions to respond honestly, along with a brief scenario explaining the benefits of undergraduate research participation. The remainder of the Control packet included other non-ADHD related information (i.e., reasons to participate in research, and the Academic Integrity Code). The Malingering group was provided with simulation instructions and a brief scenario describing a person who might "fake" the diagnosis of ADHD. The remainder of the Malingering packet included information about ADHD symptoms and diagnostic/evaluation processes adapted from the WebMD ADD and ADHD Health Center website. All participants were informed that those who adhered to the group assignment, either by responding honestly or

faking ADHD without detection (i.e., extreme responses on all items) would be entered into a raffle drawing for a $100 gift card.

Participants had five minutes to review the group instructions and corresponding information. After time elapsed, participants were asked to indicate the group assignment number on the online survey (Control = 0, Malingering = 3). One researcher collected the packets, and a second researcher who was blind to the study conditions administered the test battery (WMT, MARS, and CAT-A). After completing the measures, the Malingering group was instructed via the online survey to stop simulating ADHD and to complete the exit survey honestly. At the end of the study, non-ADHD participants received the debriefing letter to conclude study participation. Completion time varied across participants, but most finished the study in approximately 45-50 minutes.

**Incentives.** Incentives were offered to each participant to increase motivation and effort. The non-ADHD and ADHD participants who signed up to complete the study through SONA Systems received 1 credit hour for their psychology coursework requirement. ADHD participants recruited through the community received $40 cash for completing all study materials. All participants were entered into the raffle drawing of $100 Visa gift card for completing the study materials.

**Procedural integrity**. Researcher adherence to the ADHD and non-ADHD protocols were verified with a procedural script. Two trained researchers conducted most research sessions per a specified set of procedures and instructions. A third research assistant was present for 15 sessions and recorded the adherence to all procedures based on a procedural script. Adherence to the procedural script was calculated by dividing the number of steps completed by the total number of steps, multiplied by 100%. Adherence to the procedural script was found to be 100% across all included sessions.

# Results

## Data Preparation

**Data input and consistency checks.** The majority of data was collected via Qualtrics online survey system and downloaded into an Excel spreadsheet. The Word Memory Test software produced an independent data output that was entered into the participant's online survey by the primary researcher or trained research assistant. Data in Excel were then transferred to IBM SPSS Statistics software for data analyses.

**Data inspection and inclusion/exclusion criteria.** All data were visually inspected and discarded if less than 95% of data across all measures were completed. Next, data were examined in order to preserve the quality of the groups. Data were removed if a participant indicated lack of adherence or effort for the assigned condition (honest or malinger) on the final exit survey. Furthermore, participant data with clear evidence of unreliable/invalid performance (i.e., same answer for all items) and data with at least one catch item failure (i.e. "respond 3 if you are still reading this survey") were discarded due to poor study effort.

Because this study relied upon self-report of ADHD diagnosis, an external ADHD rating scale (CAT-A clinical scales) was used to confirm presence/absence of ADHD symptoms. To preserve the quality of non-ADHD group membership of the Control group, participants that reported elevated symptoms (T score $\geq$ 60) on the CAT-A Current Symptom Clinical Index, and at least one CAT-A Current symptom scale (Inattention, Impulsivity, Hyperactivity) were removed from analyses. Because an individual most likely would not be considered for an ADHD diagnosis in the absence of elevated symptom reports, participants in the ADHD and Malingering groups were removed if individuals failed to report elevated symptoms (T score $\geq$ 60) on the Current Symptom Clinical Index and at least one Current symptom scale (Inattention, Impulsivity, Hyperactivity).

**Assessing assumptions.** In preparation for the omnibus tests, data were assessed for outliers, skewness, kurtosis, homogeneity of variance, and covariates. First, the measures were examined for outliers, and normality for each index was examined by Q-Q plots, histograms, skewness, and kurtosis. For the MARS indices and the CAT-A scales, there were no instances of outliers, and skewness and kurtosis fell within acceptable ranges (< 1.5; George & Mallery, 2009). However, the WMT data contained outliers and had skewness > - 2.0. Furthermore, Levene's *F* tests revealed the assumption of homogeneity of variance was violated for the WMT, CAT-A Infrequency item count, and the SV-Index ($p < .001$). Because these validity tests were designed to produce disparate outcomes between honest reporters and those suspected of malingering (especially those instructed to fake), the unequal variances between the groups on the WMT were expected and not uncommon in malingering research (e.g., Jasinski et al., 2011). Nonetheless, further exploratory analyses were conducted using non-parametric analyses to determine whether the violation of the ANOVA assumptions altered the final results. The non-parametric equivalent of the ANOVA, Kruskal-Wallis one-way analyses of variance were conducted with the SV-index, the WMT subtests, and WMT calculation. Overall, the non-parametric analyses yielded similar findings as the original parametric ANOVA tests, thus only parametric analyses are reported. Lastly, age and sex were not found to be significant covariates for all primary measures.

**Detection of Malingering**

Between-groups comparisons were assessed with one-way analyses of variance (ANOVA) on the primary measures. Bonferroni corrections were utilized to control for the effect of repeated contrasts (10 total), resulting in an alpha level of .005. Due to unequal variances in the validity tests, Games-Howell post-hoc tests were used to examine pairwise comparisons across the groups. Effect sizes for the ANOVAs were calculated using eta-squared ($\eta^2$), with

small effect size $\eta^2 = .01$, medium effect size $\eta^2 = .06$, and a large effect size $\eta^2 = .14$. Furthermore, Cohen's *d* was calculated to present effect sizes for the pairwise comparisons across all primary measures, with small effect size *d* = .20, medium effect size *d* = .50, and a large effect size *d* = .80 (Cohen, 1988).

Classification accuracy calculations were conducted to determine the utility of each validity test to classify cases of Malingering between Malingering and ADHD groups. Using the manual or research recommended cut score, crosstab analyses were used to identify the confusion matrix (true positive, true negative, false positive, false negative classifications). Classification accuracy rates were derived from the confusion matrix, including correct classification accuracy, sensitivity, specificity, and predictive power. Sensitivity indicates the percent of Malingering participants correctly classified by a positive test outcome (fail/noncredible), and specificity reflects the percent of ADHD participants correctly classified as "honest" by a negative test outcome (pass/credible). Predictive power estimates reflect the performance of the test to accurately classify an individual within a given population. These rates account for base rate of the population, thus both positive (PPP) and negative predictive power (NPP) estimates were calculated from the base rate of simulated malingering in the study's classification accuracy analyses (Malingering = 60%). Predictive power estimates were also calculated with a base rate of malingering (25%). This base rate has been used in six previous archival studies on feigned ADHD (Harrison & Edwards, 2010; Hirsch & Christiansen, 2015; Marshall et al., 2010, 2016; Suhr et al., 2008; Suhr, Sullivan, et al., 2011).

Receiver operating characteristic (ROC) analyses were conducted to examine additional characteristics of the validity tests, including Area Under the Curve (AUC). For the ROC analyses of the MARS and CAT-A, the predictor test was the total score from the measure, with the highest value for each test predicting cases of Malingering. The ROC analyses with the

WMT had the lowest percent score from each measure predicting cases of Malingering. The AUC values range from .50 to 1.00, and can be classified as fail (.50 - .60), poor (.60 - .70), fair (.70 - .80), good (.80 - .90), and excellent (Swets, 1986).

**Validation of the original SV-index 7.** The first aim of the study was to validate the original SV-index's classification accuracy to detect malingering. Before analyzing the classification accuracy of the 7-item SV-index (SV-index 7), a one-way ANOVA was conducted on the index to determine differences between the three groups, with post-hoc tests utilized to assess pairwise comparisons. It was expected that significant differences would be found between all groups, following a pattern of Malingering group scores > ADHD group scores > Control group scores.

Group means, standard deviations, and effect sizes are reported in Table 3. As expected, there was a significant main effect across the groups on the SV-index 7, $F$ (2, 147) = 134.65, $p$ < .001, $\eta^2$ = .65. Consistent with the expected pattern, the Malingering group reported significantly higher scores ($M$ = 29.63, $SD$ = 11.71) compared to the ADHD group ($M$ = 15.29, $SD$ = 7.80); and both ADHD and Malingering groups reported statistically higher scores than the Control group ($M$ = 3.27, $SD$ = 3.63). Large effect sizes were also found between ADHD and Malingering groups ($d$ = 1.40) and between ADHD and Control groups ($d$ = 2.13).

Next, a ROC analysis was conducted to determine the ability of the SV-index 7 to discriminate cases of Malingering from clinical ADHD. Using the cut score identified to favor specificity, this study assigned a dichotomous outcome of a positive/noncredible test outcome (total score ≥ 28.0) and a negative/credible test outcome (total score < 28.0). These cut scores were derived from the Potts (2016) dataset. Next, crosstab analyses were used to identify the confusion matrix to calculate classification accuracy rates to differentiate Malingering from

ADHD. It was expected that the SV-index 7 would maintain high classification accuracy to detect Malingering participants from those with the clinical disorder.

The AUC, standard error, and confidence intervals of the SV-index 7 are presented in Table 4, and classification accuracy calculations for the index to discriminate cases of malingering can be found in Table 5. The ROC analysis found that the SV-index 7 was a good discriminator of Malingering and ADHD (AUC = .85). Classification accuracy calculations also indicated that the SV-index 7 with a cut score ≥ 28.0 had high specificity (94.7%), reflecting that this measure continues to display excellent abilities to not misclassify honest reporting clinical cases as malingering. However, the present study found lower sensitivity (56.1%) to detect cases of malingering. The index also demonstrated good overall classification accuracy to differentiate Malingering and ADHD participants (71.6%). While this classification accuracy was lower than expected, the current results still reflect that the SV-index 7 could demonstrate moderate effectiveness at malingering detection.

**Creation of a revised MARS validity index.** The original 7-item SV-index did not perform as well as it did in the original study (Potts, 2016). Therefore, the study aimed to revise the SV-index by retaining items with good detection abilities, discarding ineffective original items, and finding items from the pool that appeared to be better discriminators of simulated malingering.

To inform a decision regarding which SV-items to retain/remove, a binary logistic regression was used to identify the ability of each SV-item to predict cases of Malingering. The dependent variables were the seven SV-items' scores, and the predictor variable was Malingering (Malingering = 1, ADHD = 0). Results from the regression analyses found that two SV-items were the best predictors of Malingering, including SV Item #2 ($\beta$ = -.39, $p$ = .04), and SV Item #3 ($\beta$ = .49, $p$ = .005). Three items were not significant predictors of Malingering,

including SV Item #7 (β = .16, $p$ = .30), SV Item #6 (β = .10, $p$ = .69), and SV Item #5 (β = .21, $p$ = .36). Two other SV-items approached significance but failed to meet traditional alpha levels (< .05), SV Item #4 (β = .30, $p$ = .08), and SV Item #1 (β = .26, $p$ = .09). Given that these items appeared moderately effective in the current sample, and also performed well in prior research (Potts, 2016), these two items were considered appropriate to retain within a revised SV-index for further analyses.

After three items were removed from the original SV-index 7, alternative items for a revised SV-index were identified from the large SV item pool contained in both the current data set and previous data set (Potts, 2016). The selection of alternative SV items was based on the following metric: items had to be infrequently endorsed in the ADHD group ($M \leq 2.0$) yet frequently endorsed in the Malingering group ($M \geq 5$) and have a large mean difference between the groups ($M \geq 3.5$). This methodology identified six additional items that were not on the SV-index 7. Only two of these six items met the same infrequency and mean difference criteria across both data sets, suggesting some ability to differentiate ADHD from Malingering cases, and therefore, worthy of inclusion in a revised index. This index revision process created a new 6-item SV-index (SV-index 6) comprised of the four best predictors from SV-Index 7 and the two new SV-items identified through further analyses of two data sets.

Next, ROC analyses were conducted on the Potts (2016) data set. The intent was to identify a cut score for the SV-index 6 that provided best overall sensitivity, specificity and classification accuracy. These analyses identified a cut score ≥ 22.5 on the SV-index 6, which resulted in specificity of 87.2%, sensitivity of 91.1%, and correct classification accuracy of 89.5%. These classification accuracy calculations suggest that a score ≥ 22.5 on the SV-index 6 could have improved detection of malingering, especially when compared to the ability of the SV-index 7 in the same data set (sensitivity = 78.6%, specificity = 89.7%, accuracy = 83,2%).

Yet, these classification accuracy calculations are most likely inflated by the use of the same data set to set the cut score and to calculate classification rates. Thus, the next step was to validate the revised SV-index 6 in a different sample.

**Validation of the revised SV-index 6.** A one-way ANOVA was conducted to compare SV-index 6 mean scores across all three groups. Games-Howell post-hoc tests were conducted to examine the significance of differences for pairwise comparisons. It was expected that the analyses would follow a similar pattern as the original SV-index 7, in that the Malingering group would demonstrate significantly higher scores compared to the ADHD group, and the ADHD group would obtain significantly higher scores than the Control group.

Means, standard deviations, and effect sizes are listed in Table 3. As expected, there was a significant main effect between the groups on the revised SV-index, $F$ (2, 147) = 170.42, $p <$ .001. The revised SV-index 6 had a large effect size for the main effect ($\eta^2 = .70$), and large effect sizes between ADHD and Malingering group ($d = 1.61$), and ADHD and Control groups ($d = 2.21$). Following a similar pattern as the SV-index 7, the Malingering group reported significantly higher mean scores ($M = 27.54$, $SD = 9.35$) compared to the ADHD group ($M =$ 14.18, $SD =6.67$), and both of these groups had significant higher scores than the Control group ($M = 3.11$, $SD =3.56$).

Next, a ROC analysis was conducted to identify the ability of the SV-index 6 to differentiate Malingering and ADHD participants. Crosstab analyses and classification accuracy calculations were conducted using the cut score ($\geq 22.5$) identified on the 2016 data set. The ROC analysis for the SV-index 6 is found in Table 4, and classification accuracy calculations can be found in Table 5.

Analyses demonstrated that the revised SV-index 6 was more effective than the original SV-index 7 to detect cases of malingering. The ROC analysis determined that the revised SV-

index 6 was a good discriminator of malingering (AUC = .87). At a cut score of ≥ 22.5, the SV-Index 6 demonstrated higher classification accuracy (80.0%) compared to the SV-index 7 (71.6%). The higher accuracy rate can be attributed to the SV-index 6's higher sensitivity (75.4%) to detect malingering compared to the original index (56.1%). But as expected, such high sensitivity of the SV-index 6 resulted in slightly less than optimal specificity (86.8%) for honest reporting ADHD participants. Although the original SV-index 7 did not perform as well as expected in the current sample, the revised SV-index 6 improved detection of malingering and resulted in higher classification accuracy.

**MARS Symptom and FI over-report indicators.** Next, analyses were conducted to examine the ability of the over-report indicators, or higher cut scores on the MARS clinical indices (I-index, HI-index, Total Symptom index, and FI-index) to detect malingering. First, four separate one-way ANOVAs were conducted to assess differences between the three groups on each MARS index, and Games-Howell post-hoc tests were utilized to assess significance of pairwise comparisons. It was expected that the Malingering group would report significantly higher scores on these indices compared to the ADHD group.

Group means, standard deviations, and effect sizes for the MARS indices are presented in Table 3. The ANOVAs indicated statistically significant main effects across the MARS symptom indices and FI-index, including Total Symptom Index $F$ (2, 147) = 277.79, $p < .001$; I-Index, $F$ (2, 147) = 211.68, $p < .001$; HI-Index, $F$ (2, 147) = 236.34, $p < .001$; and FI-Index, $F$ (2, 147) = 160.54, $p < .001$. Large effect sizes were also obtained for the main effects, ranging from $\eta^2$ = .69 - .79. In a focused discussion on the pairwise comparisons between ADHD and Malingering groups (ADHD and non-ADHD group comparisons to be discussed in the relevant section), the Malingering group had a significantly higher mean score in comparison to the ADHD and

Controls groups across the MARS symptom and impairment indices. Large effect sizes were obtained between Malingering and ADHD groups across all four indices ($d$ = .72 - 1.36).

Next, the utility of the MARS over-report indicators to discriminate cases of malingering was determined with ROC analyses and classification accuracy calculations. Crosstab analyses and subsequent classification accuracy calculations were conducted with cut scores found by Potts (2016) to detect the over-report response style of malingering at ~90% specificity of ADHD: MARS Total Symptom index score ≥ 100.5, MARS HI-index score ≥ 49.5, MARS I-index score ≥ 55.5, and FI-index score ≥ 127.5. Using these cut scores, the study assigned dichotomous outcomes such that a total score at or above the cut score equaled a positive test outcome (noncredible), while a score below the cut score reflected a negative test outcome (credible). Based upon prior research (Potts, 2016), it was expected that the MARS over-report indicators would demonstrate high classification accuracy (> 70%) to discriminate cases of Malingering from diagnosed ADHD.

The ROC analyses and classification accuracy calculations for the MARS over-report indicators can be found in Table 4 and Table 5, respectively. Contrary to expectations, the classification accuracy calculations suggested that the over-report indicators on MARS symptom indices may be less effective than what was established in prior research (Potts, 2016). The ROC analyses between Malingering and ADHD groups found that the Total Symptom index, HI-index, and I-index fell within fair to good AUC ranges (AUC = .68 - .79). Classification accuracy calculations indicated that the over-report indicators had low overall correct classification accuracy (50.5% to 61.1%) to discriminate cases of Malingering and ADHD. Across the symptom indices, sensitivity ranged from 28.1% - 45.6%; and specificity for ADHD (84.2% to 86.8%) fell below optimal levels for a validity test (~90%). These results do not

support use of these over-report indicators on the MARS symptom indices to accurately classify malingering from clinical cases.

Compared to the symptom indices, FI-index demonstrated a higher AUC (.82), which can be attributed in part to a high cut score that resulted in perfect specificity for ruling out ADHD participants (100%). In other words, no ADHD participants rated impairment levels above the cut score of ≥ 127.5. In fact, the ADHD group mean on the FI-index was much lower (*M* = 72.58, *SD* = 26.27) than the Potts (2016) cut score. While this over-report indicator had excellent specificity, it also had low sensitivity to classify Malingering participants (29.8%), who as a group reported a much lower impairment score (*M* = 109.09, *SD* = 27.75) than to the current threshold. In summary, there is a high probability that a positive outcome on this indicator reflects the over-report response style of Malingering. But, this over-report indicator did not have much capacity to actually detect cases of simulated malingering in the present sample.

**CAT-A Infrequency scale.** Data were used to analyze the ability of the CAT-A Infrequency scale to detect cases of simulated malingering. Between groups comparisons were analyzed on the CAT-A Infrequency scale item count (manual's recommended method to calculate Infrequency scale). Item count was calculated by adding the number of items with the highest response option (strongly agree = 1 point, maximum score = 10). Additionally, due to possible restriction of range effect that could occur with the item count, comparisons were analyzed using the total score from the scale. The total score was derived by totaling the response scale options (1 - 4) across the 10 items (total score = 40). Two separate one-way ANOVAs were conducted with the two calculation methods for the CAT-A, and follow-up Games-Howell tests were used to examine group differences. It was expected that both the CAT-A item count and total score would follow a similar pattern of Malingering > ADHD > Control

groups. It was also expected that the CAT-A would have smaller effect sizes between the groups compared to the MARS SV-indices.

Group means, standard deviations, and effect sizes are located in Table 6. As expected, significant main effects were obtained on the CAT-A Infrequency item count, $F$ (2, 147) = 25.38, $p < .001$, $\eta^2 = .26$, and Infrequency scale total score, $F$ (2, 147) = 157.31, $p < .001$, $\eta^2 = .68$. Post-hoc analyses obtained significant findings between the groups on total score, with the Malingering group ($M$ = 29.47, $SD$ =3.34) reporting a higher total score compared to the ADHD group ($M$ = 25.95, $SD$ = 3.08), and both ADHD and Malingering groups reported higher scores than the Control group ($M$ = 18.80, $SD$ = 3.19). However, in regards to the CAT-A item count, there were no significant differences in the number of items endorsed with "strongly agree" between ADHD ($M$ = 1.32, $SD$ = 1.32) and Malingering groups ($M$ = 2.02, $SD$ = 1.95), although both groups endorsed significantly more items compared to Controls ($M$ = .15, $SD$ = 46). The CAT-A item count yielded a smaller effect size ($d$ = .41) between ADHD and Malingering groups compared to the CAT-A total score method ($d$ = 1.10); but as expected, such effect sizes were smaller than those obtained on the original SV-index 7 and the revised SV-index 6.

The CAT-A Infrequency scale's utility to discriminate cases of Malingering from ADHD was subsequently determined with ROC analyses and classification accuracy calculations. ROC analyses were conducted on both item count and total score methods, with the highest value predicting cases of malingering. Classification accuracy rates were calculated using the test's manual cut score (≥ 4 items = strongly agree; Bracken & Boatwright, 2004), and the lower cut score recommended by prior research (≥ 3 items = strongly agree; Marshall et al., 2010). An item count at or above the cut score was considered a positive test outcome (noncredible) and those below the cut score reflected a negative test outcome (credible). Crosstab analyses were used to

calculate the classification accuracy rates at both cut scores. It was expected that the CAT-A Infrequency scale would be less sensitive to Malingering in comparison to the SV-indices.

ROC analyses and classification accuracy calculations for the CAT-A Infrequency scale can be found in Tables 7 and 8, respectively. The analyses supported the expectation that the CAT-A Infrequency scale would be less effective at detecting malingering than the SV-indices. ROC analyses revealed AUC values that fell within the poor range for the item count method (AUC = .60) and the fair range for the total score (AUC = .79). Classification accuracy calculations indicated that while the CAT-A item count (≥ 4 items) could rule-out cases of ADHD effectively (specificity = 94.7%), this cut score yielded weak sensitivity for malingering detection (19.3%) and an overall classification accuracy of 49.5%. Using a lower threshold (≥ 3 items) suggested by prior research (Marshall et al., 2010), resulted in a slight increase in sensitivity (36.8%), but had suboptimal specificity (78.9%), and low overall classification accuracy for both ADHD and Malingering (53.7%). Exploratory analyses of the ROC curve for the CAT-total score determined that this method was also ineffective. Specifically, a score of ≥ 30.50 reflected high specificity (95%), but low sensitivity for cases of malingering (42%). And lowering the threshold by one point (total score ≥ 29.50) produced inadequate specificity (81.6%) for a validity test. These results indicate that the CAT-A Infrequency scale, regardless of method (item count or total score) or cut score (≥ 3 or ≥ 4) was not an effective discriminator of simulated Malingering from clinical ADHD.

**Word Memory Test.** Analyses were subsequently conducted to analyze the WMT's effectiveness to differentiate Malingering from ADHD participants. Following a similar procedure as the other measures, four separate one-way ANOVAs were conducted to investigate the utility of the WMT measures (IR, DR, CNS, and MC). The dependent variables were WMT IR subtest percent correct, WMT DR subtest percent correct, WMT CNS percent score, and

WMT MC subtest percent correct. Due to unequal variance between the groups, Games-Howell tests were used to explore pairwise comparisons. It was expected that there would be significant differences between the groups on the WMT measures, but that the effect sizes for the WMT would be smaller compared to the MARS SV-indices.

Group means, standard deviations, and effect sizes can be found in Table 6. The ANOVA analyses found significant main effects for the WMT: WMT IR, $F$ (2, 147) = 11.74, $p$ < .00l; WMT DR, $F$ (2, 147) = 29.51, $p$ < .001; WMT CNS, $F$ (2, 147) = 30.41, $p$ < .001; WMT MC, $F$ (2, 147) = 30.91, $p$ < .001. For the main effects, the four WMT measures produced moderate to large effect sizes ($\eta^2$ = .13 - .30); but as expected, such effect sizes were smaller when compared to the magnitude of differences between the groups for the MARS SV-indices.

Post-hoc analyses revealed significant pairwise comparisons between all the groups on three of the four indicators (DR, CNS, and MC), with the Malingering group obtaining a significantly lower mean score compared to the ADHD group, and both groups had significantly lower scores than the Control group. However for the WMT IR, there were no significant differences between ADHD ($M$ = 95.29%, $SD$ = 5.93%) and Malingering groups ($M$ = 91.18%, $SD$ = 13.38%), although both groups were significantly lower than Controls ($M$ = 99.25%, $SD$ = 1.47%). On all four WMT measures, the mean scores for the ADHD and Control groups fell in "pass" ranges (> 82.5 %). Interestingly, the Malingering group also had an average score in the "pass" range on the WMT IR and DR subtests. On average, the Malingering group's scores only fell in the "fail" range (mean score ≤ 82.5%) on the WMT CNS ($M$ = 81.26%, $SD$ = 18.71%) and the MC subtest ($M$ = 73.77%, $SD$ = 22.64%).

Next ROC analyses and classification accuracy calculations were used to investigate whether the four WMT measures could correctly classify cases of simulated Malingering from ADHD participants. For the ROC analyses, the predictor test was the percent score from the

WMT measure, with the lowest value predicting cases of Malingering from ADHD participants. The study utilized the manual's recommended cut score to assign dichotomous outcomes of a positive test outcome ("fail" or suboptimal effort) as a score ≤ 82.5% and a negative test outcome ("pass" or good effort) as a score > 82.5% on the WMT IR, DR, and MC subtests, and the CNS calculation. Crosstab analyses and corresponding classification accuracy calculations were derived from these cut points. It was expected that the WMT would demonstrate lower classification accuracy for Malingering participants in comparison to the SV-index 6 and SV-index 7.

Results for the ROC analyses and classification accuracy calculations can be found in Tables 7 and 8, respectively. Similar to the CAT-A's Infrequency scale, the four WMT measures yielded weak classification accuracy to correctly classify simulated cases of Malingering. Although the WMT IR, DR and CNS were effective at ruling-out ADHD, with specificity 89.5 – 97.4%, these validity tests had low sensitivity (21.1 - 47.4%) to detect simulated malingerers. However, the WMT IR subtest yielded the lowest sensitivity rates (21.1%), and very low correct classification accuracy (51.6%) of ADHD and Malingering participants. The analyses found that the WMT DR subtest had slightly better classification accuracy overall (67.4%), which is related to a slightly higher, yet still relatively low sensitivity to detect cases of malingering (47.4%). Though the WMT MC subtest demonstrated higher sensitivity (56.1%) than the other three WMT measures, this subtest had less than optimal specificity (78.9%), with eight ADHD participants' scores falling in the "fail" range on the MC subtest.

Subsequently, the ROC curves were examined to determine whether an alternate cut score could increase detection of malingering. Unfortunately, review of the data indicated that sensitivity of malingering could not be increased, while maintaining optimal specificity of ADHD (~90%). For example on the WMT IR subtest, sensitivity was poor (< 30%) at most cut

points. In fact, a cut score of ≤ 96.5% on the IR was needed to achieve a marginally sensitive test (47.4%), but consequently this higher cut score resulted in poor specificity of clinical cases (52.6%). The present findings align with expectations that the WMT would be less effective than the SV-indices. Results also suggest that the WMT may be a less than optimal validity test to detect cases of malingered ADHD.

**Combined Use of Multiple Validity Tests**

**MARS weighted validity index.** The next set of analyses addressed the question of whether or not the combined use of multiple validity tests could improve detection of malingering. First, analyses examined whether malingering detection would improve with a MARS weighted validity index. The MARS weighted validity index (WV-index) was created with the revised SV-index 6 and three MARS over-report indicators (HI-index, I-index, and FI-index). The Total Symptom index was excluded as this scale uses the same items as the HI-index and I-index. Each participant received a score that reflected the number of positive/noncredible outcomes on these four indices (maximum score = 4 points), and classification accuracy calculations were conducted at each cut point (≥ 1 - 4 positive validity outcomes). It was expected that the MARS WV-index would have higher specificity to rule-out ADHD and higher correct classification accuracy than the individual ability of the SV-indices or the MARS over-report indicators.

Table 9 presents the classification accuracy calculations for the MARS WV-index. Unfortunately, results found that the created MARS WV-index was less effective to detect malingering compared to the revised SV-index 6 alone. Although a cut point ≥ 1 on the WV-index had slightly higher sensitivity (78.9%) than the revised SV-index 6 and MARS over-report indicators (sensitivity = 28.1 – 75.4%), this cut point had suboptimal specificity (73.7%) to rule-out clinical ADHD. Additionally, a cut point ≥ 2 on the WV-index equated to higher specificity

(92.1%), yet also resulted in weaker sensitivity to detect malingering (43.9%). In summary, two or more MARS validity index/indicators with positive/noncredible outcomes most likely did not reflect an honest reporter in the current sample (PPP = 89.3%). But, these rates were no better than the SV-index 6 alone (sensitivity = 75.4%; specificity = 86.8%; PPP = 89.6%). Given that the MARS over-report indicators demonstrated weak detection abilities individually, the addition of these suboptimal indicators with the SV-index 6 did not increase classification accuracy compared to the SV-index 6 alone.

**Aggregation of the MARS, WMT, and CAT-A.** Next, analyses examined the aggregated ability of multiple validity tests (i.e., validity index, subtests, indicators) from the MARS, WMT, and CAT-A to classify cases of malingering. First, analyses focused on determining the classification accuracy with an increasing number of positive test outcomes on all the validity tests: SV-index 6, I-index, HI-index, FI-index, WMT IR, WMT DR, WMT CNS, WMT MC, and CAT-A Infrequency scale (SV-index 7 excluded due to overlap with the SV-index 6). One point was assigned for each positive test, such that the total score reflected the number of validity tests with noncredible/poor effort results for each participant. Classification accuracy calculations were conducted at each cut point (≥ 1 through 9 positive test outcomes).

Classification accuracy calculations for the aggregated ability of the nine validity tests can be found in Table 10. Analyses found that one positive outcome resulted in high sensitivity for malingering (96.5%), but low specificity of honest reporting clinical cases (47.7%). In other words, there was a high probability that an honest clinical case obtained a noncredible score on one of the nine validity tests. The accuracy of correct malingering and clinical classifications was the highest with use of two or more positive outcomes in the current study's sample (accuracy = 81.1%). Positive predictive values suggested that the use of two positive outcomes had a high probability (84.2%) that such outcomes reflected malingering in the sample's base rate

(simulated malingering = 60%). However in a lower base rate of malingering (25%), at least five noncredible/invalid tests would be needed to achieve a similarly high PPP (83.2%). In other words, more positive tests may be needed to achieve higher confidence that the outcomes reflect a case of malingering (and not an honest reporter) when using these tests in samples with a lower prevalence of malingering.

Because the aforementioned analyses combined all measures, regardless of effectiveness, additional analyses were conducted to examine whether one could use fewer, better quality validity tests to detect malingering. To examine this question, I selected the most effective measures among the MARS and WMT scales, specifically, the MARS SV-index 6, MARS Total Symptoms index, MARS FI-index, and WMT DR. The CAT-A was excluded from the analyses due to poor overall accuracy rates. Classification accuracy calculations for this group of tests are located in Table 11.

Results indicated that the combined use of these four validity tests was a more effective and efficient way to classify malingering compared to the combined use of nine tests of different quality. Two positive outcomes resulted in moderate sensitivity (54.4%) for malingering at optimal specificity of ADHD (92.1%) in the current sample. Positive outcomes on three out of five validity tests resulted in high certainty that the test findings reflected a case of malingering in both the current sample (PPP = 100%) and also in the lower base rate of 25% (PPP = 100%). Thus while the aggregation of several validity tests across different measures did not necessarily improve malingering detection (sensitivity) compared to the SV-index 6 alone, the reliance on multiple validity test outcomes could increase the probability that multiple positive tests most likely reflect a case of malingering (PPP).

**Post-hoc pass/fail analyses.** Following these planned analyses, this study examined whether there would be differences in symptom and impairment reports between Malingering

participants who passed a performance validity test (WMT) and those who failed the test. Because the DR subtest was identified as the most effective WMT test, this subtest was used to identify the "pass" ($n$ = 30) and "fail" ($n$ = 27) groups in the Malingering group. The dependent variables were the total score on the SV-index 6, MARS Total Symptom index, and the FI-index. Independent samples *t*–tests were conducted with Bonferroni correction to control for the multiple comparisons ($\alpha$ = .017).

Results indicated that there were no significant differences between Malingering participants who passed and those who failed the WMT DR on the MARS Total Symptom score, $t$ (55) = 1.36, $p$ = .18, the FI-index, $t$ (55) = 0.99, $p$ = .33, and the SV-index 6, $t$ (55) = 1.99, $p$ = .05 (see Table 12). These results reflect that performance on the WMT has little relation to how one responds on ADHD symptom, impairment, and symptom validity items. These findings add further support to the notion that the WMT is not an effective validity test for the purpose of detecting feigned ADHD.

**Classification Accuracy of MARS ADHD Indices to Detect ADHD**

Another primary aim of this study was to determine the effectiveness the MARS symptom and impairment indices to detect cases of clinical ADHD from honest reporting non-ADHD controls. Analyses included between group comparisons, and classification accuracy calculations between ADHD and non-ADHD controls. For the between group comparisons, four one-way ANOVAs were used to assess differences between the groups on the MARS symptom indices and FI-index. It was expected that the ADHD group would report significantly higher symptom and impairment levels compared to the Control group.

The between group comparisons for the MARS symptom indices and FI-index were presented in reference to the MARS over-report indicators, and the results can be found in Table 3. In review, the one-way ANOVAs yielded significant main effects between all three groups

across all four indices. As expected, post-hoc tests found significant differences between the two honest reporting groups, with the ADHD group reporting a significantly higher mean score than the Control group across all four MARS indices. Furthermore, large effect sizes were obtained between ADHD and Control groups across the three MARS symptom indices ($d$ = 2.90 - 3.38) and the FI-index ($d$ = 2.10). The largest effect size was obtained for the Total Symptom index ($d$ = 3.38).

Next, the discriminative ability of the MARS symptom indices and FI-index was investigated with four separate ROC analyses between ADHD and non-ADHD Controls. For the ROC analyses, the predictor test was the total score from the measure, with the highest value predicting cases of ADHD. Because rating scales are generally designed as screening measures, this study aimed to validate the cut scores that favored sensitivity found in the prior study (Potts, 2016): I-index (total score ≥ 20.0), HI-index (total score ≥ 12.5), Total Symptom index (total score ≥ 33.5), and FI-index (total score ≥ 27.5). Scores above the cut score represented elevated symptom/impairment levels, while scores below the cut score reflected non-elevated symptoms/impairment levels. Crosstab analyses were used to identify the confusion matrix to derive sensitivity, specificity, and classification accuracy. Positive and negative predictive power was calculated using the current sample's base rate of ADHD and controls (ADHD = 40.9%), and predictive power with estimated base rates of adult ADHD in the general population (5% Kessler et al., 2006; Weyandt & DuPaul, 2006). For these analyses, sensitivity indicates the percent of participants correctly classified as ADHD by each index, and specificity reflects the percent of non-ADHD participants correctly classified as non-clinical cases.

The accuracy of the index (AUC), standard error, and confidence intervals for the MARS symptom and impairment indices to discriminate cases of ADHD from non-ADHD controls are presented in Table 13. Table 14 contains the classification accuracy calculations for detecting

clinical cases of ADHD. As expected, the ADHD symptom indices and FI-index had excellent AUC (.92 - .98), and the cut scores that yielded high sensitivity in prior research (Potts, 2016) also had high sensitivity to detect the clinical condition (> 97%) in the present sample. Because these cut scores were set to favor sensitivity to rule-in clinical cases, it was unsurprising that these scales also had lower specificity for non-clinical Controls, ranging from 74.5 – 78.2% for the symptom indices and 79.6% for the FI-index. The highest classification accuracy was found for the MARS Total Symptom index, with correct classification of 87.1%.

Next, the study investigated whether classification accuracy increased with the combined use of symptoms and impairment indices. The study utilized the Potts (2016) cut scores to assign dichotomous outcomes, with a positive outcome indicating that positive outcomes occurred on both symptom and impairment indices (I-index + FI-index, HI-index + FI-index, or Total Symptoms index + FI-index), and a negative outcome in cases reflecting below threshold scores on symptom and/or impairment indices. Classification accuracy calculations were subsequently calculated with the symptom plus impairment outcomes. It was expected that the combined use of symptom and impairment indices would increase correct classification accuracy for differentiating clinical ADHD from non-ADHD controls compared to symptoms indices alone.

As expected, classification accuracy for both ADHD and non-ADHD controls increased when the individual had above threshold responses on both symptom and impairment indices (Table 14). However this increase was marginal. The MARS Total Symptom index had accuracy of 87.1%, and this rate increased to 88.2% when the Total Symptom index was combined with the FI-index. Thus, results partially support the expectation that the joint assessment of symptoms and impairment would increase diagnostic accuracy of clinical cases, although this increase was minimal in the present study's sample.

Additional analyses were conducted to determine if higher cut scores on the symptom and impairment indices would result in higher classification accuracy and decreased false positives of non-ADHD controls. The ROC curves were analyzed to identify new cut scores that favored optimal sensitivity of ADHD in the current sample (I index ≥ 27.5, HI-index ≥ 21.5, Total Symptom ≥ 56.5, and FI-index 37.5), which were all slightly higher than those derived from the previous study (Potts, 2016). Classification accuracy calculations were subsequently re-calculated using these cut points (see Table 15). Results showed that more conservative cut scores increased accuracy across the individual indices overall, and increased classification accuracy with the combined assessment of symptoms and impairment (accuracy = 90.3% - 93.5%).

**Validity and Reliability of the MARS**

Lastly, this study aimed to obtain validity and reliability evidence for the newly created MARS symptom, impairment, and symptom validity indices. Items means and standard deviations for the MARS I-index, HI-index, and FI-index are presented in Tables 16 - 18. The means and standard deviations for the SV-index 7 and the SV-index 6 can be referenced in Tables 19 and 20, respectively.

Cronbach's alpha revealed excellent internal consistency for MARS symptom and impairment indices, I-index (α = .97), H-index (α = .95), Total Symptom index (α = .98), FI-index (α = .97), and for the revised SV-Index 6 (α = .93). The internal consistency of original 7-item SV-index was lower (α = .84) but remained within an acceptable range. As expected for an ADHD rating scale, index intercorrelations were high (*r* = .86 to .90) between the I-Index, HI-index, and FI-index. Additionally, Total Symptom index was highly correlated with the FI-index (*r* = .84).

Correlations were used to assess validity for the MARS indices. Convergent validity was analyzed with Pearson correlations between the MARS symptom indices total score and the CAT-A Current Symptom scales and Current Symptom Clinical Index T-scores.

The correlation matrix can be found in Table 21. Results found that the CAT-A and MARS assess similar domains, with the MARS I-index demonstrating a strong correlation with CAT-A Current Inattention scale ($r$ = .92), and slightly lower, albeit still high, correlations with the CAT-A Current Hyperactivity scale ($r$ = .77) and the CAT-A Current Impulsivity scale ($r$ = .81). The MARS HI-index also had a slightly higher correlation with the CAT-A Current Hyperactivity scale ($r$ = .87), followed by the CAT-A Current Impulsivity scale ($r$ = .83) and CAT-A Current Inattention scale ($r$ = .84). Lastly, high correlations were found between the MARS Total Symptom index and the CAT-A Current Symptom Clinical Index ($r$ = .92).

Pearson correlations were also used to establish evidence of convergent validity for the SV-Index 7 and the SV-Index 6. Results found that both the SV-Index 7 and SV-Index 6 were strongly related to the CAT-A Infrequency scale total score, with $r$ = .81 and $r$ = .83, respectively. Additionally, analyses found similar, but slightly lower correlations between the CAT-A Infrequency scale item count and the SV-indices ($r$ = .64). Lower correlations were obtained between both SV-indices and the WMT, with correlations ranging from $r$ = -.32 to -.40. These findings present data to support the reliability and convergent validity of the MARS symptom indices, FI-index, and the SV-indices.

## Discussion

This study examined the ability of a new comprehensive rating scale (MARS) to detect feigned ADHD, as well as differentiate clinical ADHD from non-ADHD college students. The MARS original SV-index 7 did not fully replicate prior research findings (Potts, 2016), and therefore, a revised SV-index 6 scale was formed. This index differentiated the Malingering and

ADHD groups with high sensitivity (75.4%) at close to optimal specificity (86.8%), including an overall classification accuracy of (80.0%). Neither the WMT nor CAT-A Infrequency scale performed as well as the SV-indices, with the two published tests demonstrating either sensitivity (< 50%) and/or specificity rates (< 85%) that were not acceptable for clinical practice. A combination of the validity tests from the MARS, WMT and CAT-A did not increase overall classification accuracy beyond the SV-index 6 alone. However when using a smaller group of higher quality validity tests from the MARS and WMT, multiple positive tests did increase positive predictive power, suggesting that the combined test outcomes actually reflect malingering. With regard to differentiating ADHD and non-ADHD students, the MARS indices performed quite well. The indices led to excellent sensitivity for clinical cases of ADHD (> 90%) and generally good classification accuracy rates (79.6 - 87.1%). By combining the MARS indices, classification accuracy (88.2%) increased slightly over any single index. Overall, the results suggest that the MARS is a promising rating scale for the purposes of ADHD diagnoses and detection of malingering.

**Detection of Malingering**

**Validation of SV-index 7.** A major aim of this study was to see if the original SV-index would perform as it did in Potts (2016) when applied to new samples of college students. Current results found partial support for the original validity index. Specifically, the SV-index 7 demonstrated high specificity (94.7%), moderate sensitivity (56.1%), and a correct classification accuracy of 71.6%.

While sensitivity and overall accuracy of the SV-index 7 was lower than Potts (2016), such results were partially expected. The original study on the SV-index 7 used the same sample to identify the SV-index as well as to conduct classification accuracy analyses. Consequently, the classification accuracy rates from Potts (2016) were presumed to over-estimate accuracy, and

those rates would be lower when validated in new samples. Despite slightly lower rates in the present study, the performance of the SV-index 7 is still considered to be an improvement over any existing validity measure (Tucha et al., 2015). While the original SV-index 7 appears to be a useful measure of malingering, it seemed prudent to examine which original SV-items to retain and to discard, and to determine if items from the exploratory pool might have improved abilities to detect malingering.

**Creation of revised SV-index 6.** Following this logic, this study aimed to revise the original SV-index 7. Three ineffective SV-items were replaced with two new SV-items (selected from the exploratory pool) that appeared to be better at differentiating Malingering and ADHD groups. The revised SV-index 6 performed better than the original index, with improved accuracy of 80% to differentiate Malingering from ADHD in the current sample. While there is no "gold standard" of classification accuracy for validity tests, the obtained rates for the SV-index 6 align with the rates for some of the best neurocognitive validity tests (Sollman & Berry, 2011). In other words, the SV-index 6 performs as well as those tests relied upon in clinical and forensic settings (e.g., Social Security, workers compensation) to arrive at correct malingering classifications. Thus, the original SV-index 7 was relatively effective at detecting a new group of simulators, but the revised SV-index 6 was able to improve detection of this population.

In essence, the SV indices are intended to fool malingerers to endorse items that they think reflect characteristics of ADHD, but in fact are not common in the ADHD population. To fool a malingerer, symptom validity items often tap into false perceptions, stereotypes, and misconceptions that the general population may have about the disorder (Rogers, 2008). Research has highlighted that the general public is *aware* of ADHD, but this awareness primarily extends to basic knowledge about externalizing behaviors, such as excessive movement (McLeod, 2007). Therefore, it is not surprising that the identified SV-items all reflect

externalizing behaviors of hyperactivity and impulsivity, albeit extreme and sometimes bizarre forms of these behaviors. Furthermore, research finds that non-ADHD college students tend to have more negative attitudes regarding behaviors associated with ADHD, and tend to rate those with ADHD as less socially desirable compared to typical peers (Lebowitz, 2016). Such negative attitudes could elicit a more exaggerated response style on items that reflect extreme, unusual, and negative symptoms/behaviors. While the actual motivations for the malingerers' responses are unknown at this time, it is possible that unsystematic observations and misconceptions about ADHD played a role in their responses. Further research into the response biases of ADHD malingerers could help to identify additional items or strategies to improve detection.

Although the revised SV-index 6 was effective in the present sample, the need to revise the original SV-index highlights the challenges in detecting feigned ADHD. Research has found that both verified ADHD and feigned ADHD groups are heterogeneous across a number of characteristics (Musso & Gouvier, 2014). Such heterogeneity in both groups creates a two-fold challenge. First, a symptom validity index should contain items with high face validity to the disorder, yet be endorsed infrequently by those with the disorder. Simultaneously, these items should be endorsed frequently by those malingering ADHD, even though both groups manifest different combinations of symptoms, impairment, and cognitive deficits (Musso & Gouvier, 2014). Thus, while the SV-index 6 was an effective measure with the current sample, additional research and replication is warranted to validate this index's ability to detect cases of malingering in other groups of adults.

**MARS over-report indicators.** Another aim of this study was to replicate the effectiveness of over-report indicators (i.e., higher cut scores) on the MARS symptom and impairment indices to detect malingering from ADHD. The results confirmed expectations that those instructed to malinger ADHD would endorse significantly higher levels of symptoms and

impairment than those with ADHD. While AUC values for the I-index, HI-index, Total Symptoms index, and FI-index fell in the fair to good range (.68 - .82), the overall classification accuracy of these over-report cut scores indicated poor differentiation of malingering from clinical ADHD (accuracy = 50.5 - 61.1%). Contrary to expectations, elevated symptom or impairment scores on an expanded response scale does not accurately discriminate feigned cases from clinical ones.

Multiple studies, including the present one, find that ADHD symptom and impairment scales have weak abilities to detect malingering (Marshall et al., 2010; Tucha et al., 2015). It stands to reason that individuals with true ADHD are expected to endorse these items at fairly high levels (Barkley et al., 2008; Marshall et al., 2016). On the other hand, while some malingerers endorse ADHD items at slightly higher rates than individuals with ADHD, some malingerers tend to respond more cautiously and believably (Musso & Gouvier, 2014; Tucha et al., 2015). The end result is two groups that are not far apart in endorsements, making the difference in their responding relatively undetectable. That said, it might be the case that on an individual level, an extremely high score on these scales could signal a greater likelihood of feigning. In the current study, those who had the highest scores on the MARS indices (top 10% on symptom indices and top 20% on the FI-index) were all in the Malingering group. If such a finding could be replicated and validated, it could help clinicians screen for possible over-report bias and help determine when additional validity tests should be administered.

**Comparison of validity tests.** Results from this study found that the original SV-index 7 and revised SV-index 6 demonstrated the best classification accuracy (ADHD vs. Malingering groups) compared to the other measures. Specifically, the MARS over-report indicators, the CAT-A Infrequency scale, and the WMT were all determined to be weaker standalone classifiers of malingering. Taken together, these analyses indicate that higher detection accuracy can occur

with a validity test designed empirically to detect malingered ADHD. By contrast, results from this study do not support the use of the two published validity tests (CAT-A and WMT) currently being employed by researchers and clinicians to detect feigned cases of ADHD.

***Analysis of the WMT***. The present study found that the WMT was rather ineffective at detecting simulated cases of malingered ADHD. On the positive side, the WMT IR, DR, and CNS had high specificity to rule-out cases of honest reporting; on the negative side, these measures had low to moderate sensitivity to detect malingering. The WMT MC subtest did not demonstrate optimal specificity required of effective validity tests, perhaps because this subtest is slightly more challenging compared to the easier forced-choice subtests, and consequently is less able to discriminate malingering from those with true cognitive impairments (Green, 2003; Strauss et al., 2006). Of the WMT subtests/calculation, the WMT DR subtest was the best subtest relative to the other ones (sensitivity = 47.4%; specificity = 97.4%; accuracy = 67.4%). These rates are still lower than the WMT's classification accuracy to detect malingered cognitive/memory impairments (Sollman & Berry, 2011). Current findings indicate that the effectiveness of a widely used neuropsychological validity test designed to detect other feigned disorders (e.g., brain injury) is not particularly effective at detecting feigned ADHD.

Previous research on the WMT's ability to detect those faking ADHD has produced mixed results. Across several studies, the WMT IR, DR, and CNS have shown high levels of specificity (~90%) to rule-out those who honestly report ADHD and non-ADHD. However across simulation and archival studies, sensitivity rates ranged widely (30 – 74%). There are also differences in regards to which subtest is more effective. Two studies reported higher sensitivity rates for the IR subtest (43 – 63%) compared to the DR subtest (18 - 30%; Edmundson et al., 2017; Marshall et al., 2010). In contrast, the present study found that the WMT DR subtest was more effective (sensitivity = 47.4%) than the IR subtest (sensitivity =

21.1%). While the reason for such discrepancies are largely unknown, the vast majority of evidence suggests that a substantial proportion of malingerers will be less convinced to fake bad on an easy memory test (false negatives = 26 - 82%). In summary, the WMT IR, DR, and CNS could be used to screen-out honest reporting individuals displaying good effort overall. But, results across multiple studies suggest that the WMT has low to moderate ability to actually detect those individuals faking ADHD.

The WMT is a fairly easy memory test to pass, even for those with significant memory impairments (e.g., brain injury; Green, 2003). Although working memory deficits have been associated with ADHD, such deficits are not central to the diagnostic criteria (American Psychiatric Association, 2013). Therefore, coached malingerers would not be instructed to fake bad on a memory test. Interestingly, one consistent finding across studies, including the present one, is that the WMT was insensitive to a group of simulators coached with ADHD information (18 – 47%; Edmundson et al., 2017). This finding suggests that individuals who review the diagnostic criteria are perhaps primed to those characteristics most central to the ADHD diagnosis, symptoms and impairment, and not memory performance. As such, these coached simulators may have decided not to fake bad on an easy memory measure that has little face validity to the ADHD diagnosis. In summary, these findings underscore that one should not presume that all malingering measures can detect all types of malingering, especially savvy malingerers who are prepared to fake bad skillfully.

***CAT-A Infrequency scale***. Current results did not support the individual ability of the CAT-A Infrequency scale to detect cases of malingering, regardless of the cut score or the calculation method (item count or total score). The use of the manual's recommended cut score (≥ 4 items) effectively ruled out ADHD (specificity = 94.7%), but demonstrated weak sensitivity (19.3%). Use of the lower cut point proposed by research (≥ 3 items; Marshall et al., 2010),

raised sensitivity levels (36.8%), but resulted in poor specificity (78.9%). Exploratory analyses also could not identify an adequate cut score on either item count or total score calculations with the CAT-A Infrequency items that could improve classification accuracy.

The CAT-A was less sensitive to malingering in the present study compared to previous archival research (sensitivity = 58%; Marshall et al., 2010). The discrepant findings are perhaps best explained by the differences between the study's research designs. Marshall and colleagues (2010) utilized an archival study design, in which clinical archival cases of suspected malingering were identified based upon scores on validity tests. On the other hand, the present study employed a simulation research design that instructed non-ADHD participants to perform like a person with ADHD. Simulation designs tend to have higher internal validity as there is experimental control over group assignment, but simulated cases of malingering often tend to overestimate deficits compared to suspected archival cases (Brennan & Gouvier, 2006; Rogers, 2008). The relative limitations of these studies suggest that additional may be warranted to corroborate the classification accuracy of this symptom validity test. At this time, the obtained sensitivity rates are too low and variable (19 - 58%) to indicate that this measure could be relied upon as an individual detector of those faking ADHD.

**Summary on the validity tests.** The study supports the notion that a validity test specifically designed to detect feigned ADHD would outperform validity tests designed for other populations. Across two studies, empirically derived SV-indices emerged as sensitive, specific, and accurate measures to differentiate clinical ADHD cases from feigned ADHD cases. On the other hand, the MARS over-report indicators (i.e., higher cut scores on symptom and impairment indices), along with the two published validity tests (WMT and CAT-A) were relatively ineffective at discriminating these groups. The implications from these findings suggest that the revised SV-index 6 is a better detector of feigned ADHD than any other measure in this study, or

for that matter, any other option presented by research to date (Fuermaier et al., 2017; Harrison & Armstrong, 2016; Tucha et al., 2015). This is likely due to the empirical nature of the index, which generated and included items that had high sensitivity for feigned ADHD. Another potential advantage of the SV-index 6 is its high level of face validity. Unlike the WMT, the SV-index 6 contains items that appear to relate to ADHD characteristics, yet unbeknownst to malingerers, are not widely endorsed by those with clinical ADHD. Clearly, more research is needed on the SV-index 6, especially replication with different and larger samples. Yet, the findings lend hope that clinicians could use such a measure as part of a larger battery of tests to assess validity in psychological evaluations.

**Combined use of multiple validity tests.** Another goal of this study was to investigate whether aggregating multiple validity tests could improve malingering detection. In general, results indicated that combining multiple validity tests did not necessarily improve detection of malingering (i.e., sensitivity), per se. However, the use of multiple test outcomes improved specificity and positive predictive power rates to reflect that such scores most likely is a case of simulated malingering. The present study aligns with prior research that aggregating multiple validity tests helps to reduce the risk of a false positive classification (i.e., identifying an honest reporter as faking). As the number of positive outcomes increase, so does the probability that one is correctly determining a case of malingering (Edmundson et al., 2017; Larrabee, 2012; Victor, Boone, Serpa, Buehler, & Ziegler, 2009).

While positive outcomes across multiple validity tests can increase positive predictive power, the results highlight that the quality of the aggregated validity tests mattered, not just the number of tests included. In explanation, the SV-index 6 yielded high classification accuracy (80%) as a standalone validity test. But, this high classification accuracy was reduced to 53.7% - 63.2% when this index was combined with the suboptimal MARS over-report indicators. In

other words, simply combining multiple potentially effective validity tests only served to decrease one's accuracy to correctly classify a malingerer compared to the use of one quality measure alone. The findings supports previous recommendations for clinicians to selectively include and interpret data from tests that demonstrate utility to detect the intended population. The use of measures with questionable accuracy could lead to inaccurate classifications (Heilbronner et al., 2009).

**Concluding commentary.** Validity tests have become an increasingly important part of psychodiagnostic evaluations. The call for such tests in ADHD evaluations has also strengthened, yet no specifically designed validity tests have been developed for use in ADHD assessments (Musso & Gouvier, 2014). The development and piloting of the MARS SV-index 6 have produced a measure that seems to offer clinicians a metric for detection of feigned ADHD. Though such a measure may be useful for detecting faked ADHD, scores on validity tests alone are insufficient evidence of malingering. Data from well-established and research supported validity tests should be considered within the context of a comprehensive evaluation. This would include consideration of scores on all diagnostic assessments (e.g., symptoms, impairment, cognitive testing), information in the clinical interview, as well as collateral reports (Chafetz et al., 2015; Heilbronner et al., 2009; Iverson, 2006; Rogers, Sewell, Morey, & Ulstad, 1996). Similar to other diagnostic evaluations, clinicians are recommended to consider alternative explanations for the obtained scores on all tests, including validity tests. A clinician would also attempt to document any plausible evidence that an individual was manipulating his/her performance intentionally to obtain external incentives (Chafetz et al., 2015; Heilbronner et al., 2009; Iverson, 2006). In conclusion, the SV-indices could be a helpful tool, but these tests are only one source of information to determine the overall validity of a psychodiagnostic evaluation.

## Detection of Clinical ADHD

**MARS symptom and impairment indices.** Another major aim of this study was to replicate and validate the utility of the MARS symptom and impairment indices to detect cases of clinical ADHD from non-ADHD controls. Results supported expectations that the MARS clinical indices would effectively differentiate the two honest reporting groups with the cut scores from previous research (Potts, 2016). Specifically, the present study found that the MARS indices could detect individuals with ADHD at very high sensitivity rates (97.4 – 100%). These high sensitivity rates were partially expected given that the ADHD group was composed of individuals whose diagnosis was corroborated through a screening process (i.e., elevated symptom reports). As such, the near perfect sensitivity rates suggests that the MARS clinical indices are quite effective at screening-in verified cases of ADHD.

As a consequence of such high sensitivity, the MARS clinical indices had lower specificity of non-ADHD controls (67.2% - 78.2%). These specificity rates are in line with other ADHD rating scales (e.g., CAARS; Conners et al., 1998), and overall underscore the notion that a symptoms-only assessment may increase rates of false positive diagnoses of non-ADHD individuals (DuPaul et al., 2014; Gathje et al., 2008). The findings also serve to reinforce the practice of using a comprehensive battery of tests to diagnose ADHD. The assessment of symptoms alone will likely over-predict cases of the disorder, but the use of other metrics such as childhood history, observations, comorbidity consideration, and impairment, will reduce the number of false positives and make diagnosis more accurate (Barkley et al., 2008).

***Combined assessment of symptoms and impairment.*** This study also examined whether combining the symptom and impairment indices could improve differentiation of clinical cases from non-ADHD controls. Results of this study supported research showing that joint assessment of symptoms and impairment decreases false positive cases of non-ADHD controls (Bird et al.,

1988; DuPaul et al., 2014; Gathje et al., 2008). By using more conservative cut scores on symptoms and impairment indices than did Potts (2016), this study improved classification accuracy of those with an ADHD diagnoses. In illustration, a symptoms-only assessment using the Potts (2016) cut scores demonstrated classification accuracy of 84.9 - 87.1%. Classification accuracy increased to 90.3 - 93.3% when one assessed symptoms and impairment conjointly using a more conservative threshold. The findings suggest that the MARS clinical indices can assess both ADHD symptom and impairment levels, and the combination of these metrics can improve the classification accuracy of ADHD.

**Psychometric Characteristics of the MARS**

The MARS was designed to be a comprehensive rating scale to assess ADHD symptoms, functional impairment, and response validity As part of an ongoing test development project of the MARS, an important aim was to assess reliability and validity for this measure. Reliability analyses supported that the MARS indices had acceptable internal consistency, and all items could be retained to maintain this internal consistency. Although this is an expected finding for the *DSM 5*-based ADHD symptom indices, these results provide additional evidence to support the new MARS FI-index and SV-indices.

This study obtained evidence of validity for the MARS indices. The ease to which simulators could fake ADHD well on the symptom and impairment indices indicates that the MARS indices have face validity with the diagnosis. Also, the MARS symptom indices had strong associations with a measure of similar content, the CAT-A Current Symptom scales. Evidence of convergent validity was obtained with correlations between the SV-index and another symptom validity test (CAT-A Infrequency scale total score). The MARS SV-indices also had low to moderate correlations with the performance validity test that assesses feigned memory impairment (WMT), suggesting that the validity tests are similar but assess distinct

domain areas. In summary, this study presents some evidence of validity and reliability for the MARS, and suggests that this measure may have utility in the detection of clinical cases of ADHD and cases of malingering.

**Limitations**

The primary limitations of this study are related to the use of the simulation design. Although a simulation study typically has good internal validity, given experimental control of the malingering group assignment, simulation studies have limited external validity, as contrived studies cannot adequately recreate the real-world incentives and motivations for people to malinger a diagnosis. Although this study utilized a modest incentive to increase effort, a gift card is a less powerful reinforcer compared to the larger incentives for some individuals to effectively malinger ADHD (e.g., test accommodations on a high stakes exam). This study also used coaching instructions, a method informed by previous research (Rogers, 2008; Rogers & Gillard, 2011). While there is some consistency in coaching instructions across simulation studies (i.e., ADHD diagnostic information), there are inconsistencies as well. For example, some studies provide more diagnostic information, allow more preparation time, and caution simulators to avoid extreme responding. Simulation study instructions also focus on external incentives to malinger (i.e., accommodations, medication), which may not reflect the response style of those who are motivated to obtain the diagnosis for internal reasons (e.g., explanation for personal failures). As such, the manipulation used in the current study may not generalize to other simulation studies, let alone reflect all types and motivations for feigning a disorder (Brennan & Gouvier, 2006; Rogers, 2008; Rogers & Gillard, 2011). The relative strengths and weaknesses of simulation designs should be considered in the context of other research studies on feigned ADHD.

Another limitation that affects the external validity of the findings involves the characteristics of the sample. This study used a convenient sample predominantly comprised of Caucasian students between 18-20 years of age who attend a private institution in the Northeast. Although gender was not found to be a significant covariate, the groups differed slightly in regards to gender distribution, with the non-ADHD group having more females compared to the ADHD group. Furthermore, the study did not assess for pre-existing knowledge of ADHD, ADHD symptoms, or cognitive abilities, which could have affected the study outcomes. As a result, the results from the present study may not generalize to other groups.

The study is limited by the use of self-report to determine ADHD or non-ADHD status. Despite attempts to ensure presence of ADHD, it is possible that a portion of the ADHD sample included individuals that do not meet all *DSM-5* criteria for ADHD diagnosis, and also possible that some individuals with undiagnosed ADHD were included in the non-ADHD control group. On the other hand, this study may have removed individuals with clinical ADHD who may have under-reported symptoms. The selection procedures did not include clinical interviews, collateral symptom reports, or neuropsychological testing, and did not review all *DSM-5* criteria for the diagnosis. Similarly, the Malingering group only included participants who could successfully fake ADHD symptoms, and so the malingering group also was restricted by the selection procedures. Consequently, study findings do not necessarily generalize to all ADHD, non-ADHD, and malingered ADHD populations.

The generalizability of this study is also limited by the testing conditions. While ADHD participants completed the study in a private testing room, non-ADHD individuals participated in small groups in a computer lab. This test setting does not replicate the individual testing environment of a psychological evaluation, although participants were spaced within the room to afford some degree of privacy. Additionally, all participants completed the validity tests

successively, and not as part of a comprehensive psychological evaluation. Consequently the findings may not accurately reflect the ability of these validity tests to assess effort when interspersed with other diagnostic measures in a comprehensive psychodiagnostic evaluation.

It is also suspected that the length of the rating scales may have contributed to fatigue effects. The entire study took approximately 45-60 minutes for individuals to complete with good effort (e.g., reading all items). Data indicated that some individuals rushed to complete the long rating scales quickly. A significant number of individuals were removed from the data analyses because they confessed they did not read/consider the items, and some stated that the study was too long. Anticipating this possibility, the study over-recruited non-ADHD participants, and these cases could be removed while maintaining a sufficient sample size. Future research on the MARS should consider such factors to improve overall study effort.

**Directions for Future Research**

While the results of the present study support the use of the MARS indices to detect ADHD and the MARS SV-indices to detect simulated cases of malingering, it is unknown whether these indices will remain effective to detect feigning across various clinical samples. Further research employing both simulation designs and archival studies with clinical cases is recommended. The strengths of both research methodologies should help to provide the best support for creating a validity test to differentiate true cases of malingering (Rogers, 2008).

Additional research is needed in larger demographic samples to validate the usefulness of the MARS in more heterogeneous groups. Although the assumption of validity tests is that they can be easily passed by most, regardless of participant characteristics (e.g., age, gender, ethnicity, cognitive impairment), it is possible that differences across studies may be attributed, even in some small part, to differences in participant characteristics. With this in mind, future research with the MARS should extend to diverse groups, including both college and working

adult populations. This research could identify different malingering response styles across groups, and if so, what measures/methods can best detect them.

The SV-indices and SV-items would benefit from additional investigation. At this time, it is unknown what variable contributed to the identification of the original SV-items and the two additional SV-items. A detailed examination of the SV-items could help to elucidate the best target words, general content, and specific situations that appear to produce differential responding between malingering and ADHD groups. Furthermore, it may be beneficial to collect feedback from simulators about malingering strategies, items targeted, and perhaps a rationale for targeting such items. This information could inform the identification, revision, or reduction of current SV-items.

Although most validity tests use dichotomous labels of "pass/credible" or "fail/noncredible," the construct of effort lies on a continuum (Heilbronner et al., 2009; Larrabee, 2012). Some rating measures (e.g., Behavior Assessment System for Children, Third Edition; Reynolds & Kamphaus, 2015) use graded classification ranges to reflect the relative risk that clinical scores are elevated (e.g., mild, moderate severe) and the threat that responses are invalid (e.g., acceptable, caution, and extreme caution). It is possible that such graded classifications could help to improve interpretation of the overall findings, and consequently may help to communicate evaluation findings more clearly and succinctly in clinical settings. A consideration of graded classification labels, and not dichotomous outcomes, with the MARS clinical indices and symptom validity indices presents an interesting avenue for future investigations.

The MARS may also benefit from the inclusion of additional subscales that could differentiate ADHD from common comorbid conditions. Given that the *DSM-5* includes a differential diagnosis criterion, the MARS may be enhanced with indices that assess for learning disability/problems, mood disorders, disruptive behaviors, and substance use problems. Not only

will this help with differential diagnosis and treatment planning, a comprehensive rating scale that includes additional clinical items could help to sufficiently mask the smaller grouping of symptom validity items after the ineffective experimental ones are removed from the scale. A comprehensive rating scale that includes ADHD symptoms and impairment items, along with differential diagnoses and malingering items, may facilitate the accurate diagnosis of ADHD from other disorders and conditions.

This study also presents additional support for the validity and reliability of the MARS as a self-report measure to detect clinical ADHD and malingering, and to screen-out non-ADHD cases. However, this is a new self-report measure that requires additional research before it would be appropriate for clinical use. Research is needed to identify an appropriate cut score on all the MARS indices. Furthermore, continued research is needed to establish additional evidence of validity (e.g., construct validity) and reliability (e.g., retest reliability) for the MARS. The acquisition of a larger sample will also allow factor analyses to be conducted to determine the actual structure of the scale. Results from this study suggest that the MARS could be useful in clinical settings. Additional research, data collection, and analyses are warranted to continue to build evidence for this newly created ADHD scale.

**Conclusions**

Adult self-referrals for ADHD evaluations are on the rise, and evidence suggests that a proportion of these individuals are malingering the disorder to obtain incentives. Currently, there is no published ADHD measure or validity test that can detect this population. Several validity tests have shown some effectiveness, yet none of them has been empirically developed or sufficiently investigated to support its effectiveness in clinical settings. The current study evaluated a comprehensive rating scale (MARS) that assessed adults on three dimensions: ADHD symptoms, functional impairment, and symptom validity or likelihood of faking ADHD.

Findings indicated that the MARS clinical indices were effective at identifying individuals with ADHD diagnoses, and inclusion of an impairment scale could be used to make more conservative and accurate diagnostic decisions regarding ADHD. The new Symptom Validity index (SV-index 6) outperformed two existing validity tests (WMT and CAT-A Infrequency scale). A future goal is to standardize the MARS and provide clinicians with norm-referenced cut scores for diagnostic decision-making. The MARS clinical and validity indices could improve the accuracy of ADHD diagnosis within a comprehensive ADHD evaluation of young adults.

Table 1

*CAT-A Current Symptom Scales, Current Index, and Clinical Index Classification Ranges*

| CAT-A Scale/Index | ADHD (*n* = 38) | | Malingering (*n* = 57) | | Control (*n* = 55) | |
|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *n* | % |
| CAT-A Current Inattention scale | | | | | | |
| Normal range | 3 | 7.9 | 2 | 3.5 | 54 | 98.2 |
| Mild range | 20 | 52.6 | 18 | 31.6 | 1 | 1.7 |
| Significant range | 10 | 26.3 | 24 | 42.1 | 0 | 0 |
| Very Significant range | 5 | 13.2 | 13 | 22.8 | 0 | 0 |
| CAT-A Current Impulsivity scale | | | | | | |
| Normal range | 7 | 18.4 | 3 | 5.3 | 50 | 90.9 |
| Mild range | 21 | 55.3 | 23 | 40.4 | 5 | 9.1 |
| Significant range | 8 | 21.1 | 20 | 35.1 | 0 | 0 |
| Very Significant range | 2 | 5.3 | 11 | 19.3 | 0 | 0 |
| CAT-A Current Hyperactivity scale | | | | | | |
| Normal range | 13 | 34.2 | 3 | 5.3 | 52 | 94.5 |
| Mild range | 13 | 34.2 | 19 | 33.3 | 3 | 5.5 |
| Significant range | 8 | 21.1 | 22 | 38.6 | 0 | 0 |
| Very Significant range | 5 | 13.2 | 13 | 22.8 | 0 | 0 |
| CAT-A Current Symptoms Clinical index | | | | | | |
| Normal range | 3 | 7.9 | 0 | 0 | 55 | 100.0 |
| Mild range | 16 | 44.7 | 12 | 21.1 | 0 | 0 |
| Significant range | 15 | 39.5 | 24 | 42.1 | 0 | 0 |
| Very Significant range | 4 | 10.5 | 21 | 36.8 | 0 | 0 |
| CAT-A Clinical Index | | | | | | |
| Normal range | 0 | 0 | 0 | 0 | 55 | 100.0 |
| Mild range | 23 | 60.5 | 13 | 22.8 | 0 | 0 |
| Significant range | 8 | 21.1 | 24 | 42.1 | 0 | 0 |
| Very Significant range | 7 | 18.4 | 20 | 35.1 | 0 | 0 |

*Note.* The CAT-A Clinical Index includes items from both Childhood and Current Symptoms.

Table 2

*Demographic Characteristics of Final Sample*

| Characteristic | ADHD (*n* = 38) | | Malingering (*n* = 57) | | Control (*n* = 55) | |
|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *n* | % |
| Gender[a] | | | | | | |
| Males | 20 | 52.6 | 19 | 33.3 | 20 | 36.4 |
| Females | 18 | 47.4 | 38 | 66.7 | 35 | 63.6 |
| Ethnicity[a] | | | | | | |
| Caucasian/White | 34 | 89.5 | 39 | 68.4 | 41 | 74.5 |
| African American/Black | 0 | 0 | 2 | 3.5 | 3 | 5.5 |
| Hispanic | 0 | 0 | 2 | 3.5 | 0 | 0 |
| Asian | 2 | 5.3 | 8 | 14.0 | 8 | 14.5 |
| Multi-racial | 2 | 5.3 | 6 | 10.5 | 2 | 3.6 |
| Not reported | 0 | 0 | 0 | 0 | 1 | 1.8 |
| First Language[a] | | | | | | |
| English | 37 | 97.4 | 53 | 93.0 | 50 | 90.9 |
| Other | 1 | 2.6 | 4 | 7.0 | 5 | 8.3 |
| Year in School[a] | | | | | | |
| Freshman | 28 | 73.7 | 45 | 78.9 | 36 | 66.5 |
| Sophomore | 6 | 15.8 | 8 | 14.0 | 14 | 25.5 |
| Junior | 2 | 5.3 | 4 | 7.0 | 4 | 7.3 |
| Senior | 1 | 2.6 | 0 | 0 | 0 | 0 |
| Not a student | 1 | 2.6 | 0 | 0 | 0 | 0 |
| Not reported | 0 | 0 | 0 | 0 | 1 | 1.8 |
| School problems[b] | | | | | | |
| No | 21 | 55.3 | 52 | 91.2 | 54 | 98.2 |
| Yes | 17 | 44.7 | 5 | 8.8 | 1 | 1.8 |
| Academic accommodations[b] | | | | | | |
| No | 10 | 26.3 | 54 | 94.7 | 52 | 94.5 |
| Yes | 28 | 73.7 | 3 | 5.3 | 3 | 5.5 |
| Regular stimulant use | | | | | | |
| No | 7 | 18.4 | - | - | - | - |
| Yes | 31 | 81.6 | - | - | - | - |
| Stimulant use in past 12 hours | | | | | | |
| No | 20 | 52.6 | - | - | - | - |
| Yes | 16 | 42.1 | - | - | - | - |

Table 3

*Group Comparisons of MARS Indices*

| | ADHD ($n$ = 38) | | Malingering ($n$ = 57) | | Control ($n$ = 55) | | | Cohen's $d$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Index | *M* | *SD* | *M* | *SD* | *M* | *SD* | $\eta^2$ | ADHD vs. Mal | ADHD vs. C | Mal vs. C |
| Symptom Validity Index 7 | 15.29 | 7.80 | 29.63 | 11.71 | 3.27 | 3.63 | .65 | 1.40 | 2.13 | 3.05 |
| Symptom Validity Index 6 | 14.18 | 6.67 | 27.54 | 9.35 | 3.11 | 3.56 | .70 | 1.61 | 2.21 | 3.46 |
| Inattention Index | 43.97 | 11.67 | 51.47 | 9.68 | 11.71 | 10.97 | .74 | .72 | 2.90 | 3.86 |
| Hyperactivity-Impulsivity Index | 37.05 | 11.28 | 47.75 | 10.05 | 8.89 | 7.94 | .76 | 1.02 | 3.02 | 4.32 |
| Total Symptom Index | 81.03 | 18.85 | 99.23 | 18.51 | 20.60 | 17.53 | .79 | .99 | 3.38 | 4.40 |
| Functional Impairment Index | 72.58 | 26.27 | 109.09 | 27.75 | 24.65 | 20.65 | .69 | 1.36 | 2.10 | 3.47 |

*Note.* $p < .001$ for all main effects and pairwise comparisons. Mal = Malingering; C = Control.

Table 4

*Area Under the Curve from Receiver Operating Characteristic Analyses for the MARS Indices to Classify Malingering between ADHD and Malingering Groups*

| | | | Confidence Interval | |
|---|---|---|---|---|
| Index | AUC | Standard Error | Lower | Upper |
| Symptom Validity Index 7 | .85 | .04 | .78 | .93 |
| Symptom Validity Index 6 | .87 | .03 | .80 | .94 |
| Inattention Index | .68 | .06 | .57 | .79 |
| Hyperactivity-Impulsivity Index | .75 | .05 | .65 | .85 |
| Total Symptom Index | .75 | .05 | .65 | .85 |
| Functional Impairment Index | .82 | .04 | .74 | .91 |

*Note*. $p < .001$ for all indices.

Table 5

*Classification Accuracy for the MARS Indices to Detect Malingering from ADHD*

| | | | | | Current Study base rate (60%) | | Estimated base rate (25%) | |
|---|---|---|---|---|---|---|---|---|
| Measure | Cut Score | Sensitivity | Specificity | Accuracy | PPP | NPP | PPP | NPP |
| Symptom Validity Index 7 | ≥ 28.0 | 56.1 | 94.7 | 71.6 | 94.1 | 59.0 | 77.9 | 86.6 |
| Symptom Validity Index 6 | ≥ 22.5 | 75.4 | 86.8 | 80.0 | 89.6 | 70.2 | 65.6 | 91.4 |
| Inattention Index | ≥ 55.5 | 28.1 | 84.2 | 50.5 | 72.7 | 43.8 | 37.2 | 77.8 |
| Hyperactivity-Impulsivity Index | ≥ 49.5 | 43.9 | 86.8 | 61.1 | 83.3 | 50.8 | 52.6 | 82.3 |
| Total Symptom Index | ≥ 100.5 | 45.6 | 84.2 | 61.1 | 81.3 | 50.8 | 49.0 | 82.3 |
| Functional Impairment Index | ≥ 127.5 | 29.8 | 100.0 | 57.9 | 100.0 | 48.7 | 100.0 | 81.0 |

*Note.* PPP = positive predictive power, NPP = negative predictive power. Cut scores are based on Potts (2016) data analyses.

Table 6

*Group Comparisons of the CAT-A Infrequency Scale and WMT*

| | ADHD ($n$ = 38) | | Malingering ($n$ = 57) | | Control ($n$ = 55) | | | Cohen's *d* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Index | *M* | *SD* | *M* | *SD* | *M* | *SD* | $\eta^2$ | ADHD vs. Mal | ADHD vs. C | Mal vs. C |
| CAT-A Infrequency Scale | | | | | | | | | | |
| Item count | 1.32 | 1.32 | 2.02 | 1.95 | .15 | .45 | .26 | .41 | 1.29 | 1.32 |
| Total score | 25.95 | 3.08 | 29.47 | 3.34 | 18.80 | 3.19 | .68 | 1.10 | 2.29 | 3.30 |
| WMT | | | | | | | | | | |
| Immediate Recognition | 95.29 | 5.93 | 91.18 | 13.38 | 99.25 | 1.47 | .13 | .38 | 1.01 | .85 |
| Delayed Recognition | 96.61 | 5.94 | 83.12 | 18.70 | 99.58 | 0.98 | .29 | .91 | .78 | 1.24 |
| Consistency | 93.87 | 7.66 | 81.26 | 18.71 | 98.85 | 2.15 | .29 | .83 | .98 | 1.32 |
| Multiple Choice | 89.08 | 13.14 | 73.77 | 22.64 | 96.73 | 4.43 | .30 | .80 | .85 | 1.41 |

*Note.* Significant main effects for all measures $p < .001$. Post hoc comparisons indicated no differences between ADHD and Malingering on WMT IR and CAT-A Item count. Mal = Malingering; C = Control.

Table 7

*Area Under the Curve from Receiver Operating Characteristic Analyses for the CAT-A Infrequency Scale and the WMT to Classify Malingering between ADHD and Malingering Groups*

| Index | AUC | Standard Error | *p* value | Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| CAT-A Infrequency Scale | | | | | |
| Item count | .60 | .06 | .12 | .48 | .71 |
| Total score | .79 | .05 | <.001 | .69 | .88 |
| WMT | | | | | |
| Immediate Recognition | .54 | .06 | .55 | .42 | .65 |
| Delayed Recognition | .68 | .05 | <.001 | .58 | .79 |
| Consistency | .68 | .05 | <.001 | .57 | .78 |
| Multiple Choice | .68 | .05 | <.001 | .57 | .79 |

Table 8

*Classification Accuracy of the CAT-A Infrequency Scale and WMT to Detect Malingering from ADHD*

| | | | | | Current Study base rate (60%) | | Estimated base rate (25%) | |
|---|---|---|---|---|---|---|---|---|
| Measure | Cut Score | Sensitivity | Specificity | Accuracy | PPP | NPP | PPP | NPP |
| CAT-A Infrequency Scale | | | | | | | | |
| Marshall et al. 2010 | ≥ 3 | 36.8 | 78.9 | 53.7 | 72.4 | 45.5 | 36.8 | 78.9 |
| Test manual | ≥ 4 | 19.3 | 94.7 | 49.5 | 84.6 | 43.9 | 54.8 | 77.9 |
| WMT | | | | | | | | |
| Immediate Recognition | ≤ 82.5 | 21.1 | 97.4 | 51.6 | 92.3 | 45.1 | 73.0 | 78.7 |
| Delayed Recognition | ≤ 82.5 | 47.4 | 97.4 | 67.4 | 96.4 | 55.2 | 85.9 | 84.7 |
| Consistency | ≤ 82.5 | 47.4 | 89.5 | 64.2 | 87.1 | 53.1 | 60.1 | 83.6 |
| Multiple Choice | ≤ 82.5 | 56.1 | 78.9 | 65.3 | 80.0 | 54.5 | 47.0 | 84.4 |

Table 9

*Classification Accuracy for the MARS Weighted Validity Index to Detect Malingering from ADHD*

| Number of Positive/ Noncredible Outcomes | Sensitivity | Specificity | Accuracy | Study base rate (60%) | | Estimated base rate (25%) | |
|---|---|---|---|---|---|---|---|
| | | | | PPP | NPP | PPP | NPP |
| ≥1 Noncredible Outcome | 78.9 | 73.7 | 76.8 | 81.8 | 70.0 | 50.0 | 91.3 |
| ≥2 Noncredible Outcomes | 43.9 | 92.1 | 63.2 | 89.3 | 52.2 | 64.9 | 83.1 |
| ≥3 Noncredible Outcomes | 31.6 | 92.1 | 55.8 | 85.7 | 47.3 | 57.1 | 80.2 |
| ≥4 Noncredible Outcomes | 22.8 | 100.0 | 53.7 | 100.0 | 46.3 | 100.0 | 79.5 |

*Note.* MARS weighted validity index combines the SV-index 6, I-index, HI-index, and FI-index.

Table 10

*Classification Accuracy of Multiple Validity Tests to Detect Malingering from ADHD*

| | | | | Study base rate (60%) | | Estimated base rate (25%) | |
|---|---|---|---|---|---|---|---|
| Number of Positive Tests | Sensitivity | Specificity | Accuracy | PPP | NPP | PPP | NPP |
| ≥ 1 Positive Validity Test | 96.5 | 47.4 | 76.8 | 73.3 | 90.0 | 37.9 | 97.6 |
| ≥ 2 Positive Validity Tests | 84.2 | 76.3 | 81.1 | 84.2 | 76.3 | 54.2 | 93.5 |
| ≥ 3 Positive Validity Tests | 68.4 | 86.8 | 75.8 | 88.6 | 64.7 | 63.3 | 89.2 |
| ≥ 4 Positive Validity Tests | 54.4 | 92.1 | 69.5 | 91.2 | 57.4 | 69.7 | 85.8 |
| ≥ 5 Positive Validity Tests | 38.6 | 97.4 | 62.1 | 95.7 | 51.4 | 83.2 | 82.6 |
| ≥ 6 Positive Validity Tests | 17.5 | 100.0 | 50.5 | 100.0 | 44.7 | 100.0 | 78.4 |
| ≥ 7 Positive Validity Tests | 10.5 | 100.0 | 46.3 | 100.0 | 42.7 | 100.0 | 77.0 |
| ≥ 8 Positive Validity Tests | 10.5 | 100.0 | 46.3 | 100.0 | 42.7 | 100.0 | 77.0 |
| 9 Positive Validity Tests | 5.3 | 100.0 | 43.2 | 100.0 | 41.3 | 100.0 | 76.0 |

*Note.* Validity tests included the SV-index 6, I-index, HI-index, FI-index, WMT IR, WMT DR, WMT CNS, WMT MC, and CAT-A Infrequency scale.

Table 11

*Classification Accuracy of Four Validity Tests to Detect Malingering from ADHD*

| | | | | Study base rate (60%) | | Estimated base rate (25%) | |
|---|---|---|---|---|---|---|---|
| Number of Positive Tests | Sensitivity | Specificity | Accuracy | PPP | NPP | PPP | NPP |
| ≥ 1 Positive Validity Test | 94.7 | 76.3 | 87.4 | 85.7 | 90.63 | 57.1 | 97.7 |
| ≥ 2 Positive Validity Tests | 54.4 | 92.1 | 69.5 | 91.2 | 57.4 | 69.7 | 85.8 |
| ≥ 3 Positive Validity Tests | 36.8 | 100.0 | 62.1 | 100.0 | 100.0 | 100.0 | 82.6 |
| 4 Positive Validity Tests | 12.3 | 100.0 | 47.4 | 100.0 | 43.2 | 100.0 | 77.4 |

*Note.* Validity tests included the MARS SV-index 6, MARS Total Symptom, MARS FI-index, and WMT DR

Table 12

*Comparisons of Pass/Fail Rates on the WMT DR Subtest across the Malingering Group*

| | Pass ($n$ = 30) | | Fail ($n$ = 27) | | | | |
|---|---|---|---|---|---|---|---|
| MARS Index | *M* | *SD* | *M* | *SD* | *t* | *p* | *d* |
| MARS Total Symptom Index | 102.37 | 18.50 | 95.74 | 18.22 | 1.36 | .18 | .37 |
| MARS FI-Index | 112.53 | 26.97 | 105.26 | 28.60 | .99 | .33 | .27 |
| MARS SV-Index 6 | 29.83 | 8.84 | 25.00 | 9.41 | 1.99 | .05 | .54 |

*Note.* Pass/fail rates determined using manual cut score on the WMT DR subtest.

Table 13

*Area under the Curve from Receiver Operating Characteristic Analyses for the Five MARS Indices Classifying ADHD between ADHD and Control Groups*

| | | | | Confidence Interval | |
|---|---|---|---|---|---|
| Index | AUC | Standard Error | *p* value | Lower | Upper |
| Inattention Index | .97 | .02 | .00 | .93 | 1.00 |
| Hyperactivity-Impulsivity Index | .98 | .01 | .00 | .95 | 1.00 |
| Total Symptom Index | .98 | .01 | .00 | .96 | 1.00 |
| Functional Impairment Index | .92 | .03 | .00 | .86 | .98 |

Table 14

*Classification Accuracy of MARS Symptom and Impairment Indices to Detect Clinical ADHD from Non-ADHD Controls Using Potts (2016) Cut Scores*

| | | | | | Study base rate (40.9%) | | Estimated base rate (5%) | |
|---|---|---|---|---|---|---|---|---|
| Measure | Cut score | Sensitivity | Specificity | Accuracy | PPP | NPP | PPP | NPP |
| Inattention Index | ≥ 20.0 | 97.4 | 76.4 | 84.9 | 74.0 | 97.7 | 17.8 | 99.8 |
| Hyperactivity-Impulsivity Index | ≥ 12.5 | 100.0 | 74.5 | 84.9 | 73.1 | 100.0 | 17.1 | 100.0 |
| Total Symptom Index | ≥ 33.5 | 100.0 | 78.2 | 87.1 | 76.0 | 100.0 | 19.4 | 100.0 |
| Functional Impairment Index | ≥ 27.5 | 97.4 | 67.3 | 79.6 | 67.3 | 97.4 | 13.6 | 99.8 |
| I-Index + FI-Index | - | 97.4 | 80.0 | 87.1 | 77.1 | 97.8 | 20.4 | 99.8 |
| HI-Index + FI-index | - | 97.4 | 80.0 | 87.1 | 77.8 | 97.8 | 20.4 | 99.8 |
| Total Symptom Index + FI-index | - | 97.4 | 81.8 | 88.2 | 78.7 | 97.8 | 22.0 | 99.8 |

*Note.* Combined Symptom and impairment indices with Potts (2016) cut scores.

Table 15

*Classification Accuracy of MARS Symptom and Impairment Indices to Detect Clinical ADHD from Non-ADHD Controls Using More Conservative Cut Scores*

| | | | | | Study base rate (40.9%) | | Estimated base rate (5%) | |
|---|---|---|---|---|---|---|---|---|
| Measure | Cut score | Sensitivity | Specificity | Accuracy | PPP | NPP | PPP | NPP |
| Inattention Index | ≥ 27.5 | 94.7 | 96.4 | 95.7 | 94.7 | 96.4 | 58.7 | 99.7 |
| Hyperactivity-Impulsivity Index | ≥ 21.5 | 89.5 | 92.7 | 91.4 | 89.5 | 92.7 | 39.2 | 99.4 |
| Total Symptom Index | ≥ 56.5 | 89.5 | 96.4 | 93.5 | 94.4 | 92.9 | 56.7 | 99.4 |
| Functional Impairment Index | ≥ 37.5 | 89.5 | 76.4 | 81.7 | 72.4 | 91.3 | 16.6 | 99.3 |
| I-Index + FI-Index (conservative) | - | 89.5 | 96.4 | 93.5 | 94.4 | 93.0 | 56.7 | 99.4 |
| HI-Index + FI-index (conservative) | - | 81.6 | 96.4 | 90.3 | 94.0 | 88.3 | 54.4 | 99.0 |
| Total Symptom Index + FI-index (conservative) | - | 81.6 | 96.4 | 90.3 | 94.4 | 88.3 | 54.4 | 99.0 |

*Note.* Combined Symptom and impairment indices use the following cut scores: I-index ≥ 27.5, HI-index ≥ 21.5, Total Symptom index ≥ 56.5, and FI-index 37.5.

Table 16

*Means and Standard Deviations of MARS Inattention Symptom Items*

| | ADHD ($n$ = 38) | | Malingering ($n$ = 57) | | Control ($n$ = 55) | |
|---|---|---|---|---|---|---|
| Inattention Symptom Item | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Fail to give close attention to details or make careless mistakes in my work | 5.55 | 1.64 | 6.02 | 1.50 | 1.29 | 1.64 |
| Difficulty sustaining attention in tasks or play activities | 4.71 | 1.93 | 5.67 | 1.34 | 1.24 | 1.66 |
| Do not listen when spoken to directly (mind seems elsewhere) | 4.18 | 2.25 | 5.25 | 1.70 | 1.22 | 1.76 |
| Do not follow through with instructions and fail to finish schoolwork, chores, or work duties | 4.42 | 2.39 | 5.46 | 1.92 | .73 | .99 |
| Difficulty organizing tasks and activities | 4.92 | 1.94 | 5.49 | 1.72 | 1.38 | 1.38 |
| Avoid, dislike, or reluctant to engage in tasks that require sustained mental effort | 4.61 | 1.98 | 5.74 | 1.45 | 1.44 | 1.80 |
| Lose things necessary for tasks or activities | 6.39 | 1.50 | 6.61 | 1.10 | 1.91 | 1.79 |
| Easily distracted by extraneous stimuli or irrelevant thoughts | 4.89 | 2.19 | 5.63 | 1.22 | 1.47 | 1.71 |
| Forgetful in daily activities | 4.29 | 2.00 | 5.61 | 1.57 | 1.04 | 1.41 |

Table 17

*Means and Standard Deviations of MARS Hyperactivity-Impulsivity Symptom Items*

| | ADHD ($n$ = 38) | | Malingering ($n$ = 57) | | Control ($n$ = 55) | |
|---|---|---|---|---|---|---|
| Hyperactivity Symptom Item | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Fidget in seat by squirming, tapping my hands and/or my feet | 3.79 | 2.00 | 4.82 | 1.84 | .25 | .58 |
| Leave my seat when remaining seated is expected | 3.47 | 2.33 | 5.47 | 1.69 | .87 | 1.29 |
| Run about or climb in situations where it is inappropriate | 6.18 | 1.94 | 5.98 | 1.93 | 1.65 | 2.10 |
| Unable to play or engage in leisure activities quietly | 4.61 | 2.62 | 5.82 | 1.35 | 1.67 | 1.85 |
| Constantly on the go/driven by motor | 3.66 | 2.06 | 4.91 | 1.48 | .56 | .92 |
| Talk excessively | 2.39 | 2.30 | 4.14 | 1.99 | .31 | .77 |
| Blurt out an answer before a question has been completed | 4.68 | 2.35 | 5.79 | 1.37 | 1.49 | 1.71 |
| Have difficulty waiting for my turn | 4.03 | 2.37 | 5.51 | 1.49 | 1.11 | 1.21 |
| Interrupt or intrude on others (butt into conversations or activities without permission) | 4.24 | 1.94 | 5.30 | 1.40 | .96 | 1.23 |

Table 18

*Means and Standard Deviations of MARS Functional Impairment Items*

| | ADHD ($n$ = 38) | | Malingering ($n$ = 56) | | Control ($n$ = 62) | |
|---|---|---|---|---|---|---|
| Functional Impairment Item | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| In home life with immediate family | 3.05 | 2.47 | 3.86 | 2.13 | 1.04 | 1.55 |
| In getting household chores completed | 3.58 | 2.04 | 5.44 | 1.68 | 1.07 | 1.03 |
| In managing daily activities | 4.18 | 1.67 | 5.25 | 1.73 | 1.24 | 1.40 |
| In my social interactions with strangers | 2.45 | 2.02 | 4.72 | 2.04 | 1.20 | 1.45 |
| In my work/job | 3.29 | 2.17 | 5.63 | 1.63 | 1.05 | 1.45 |
| In budgeting my money, bills, and/or debt | 4.00 | 2.25 | 5.68 | 1.96 | 1.78 | 2.05 |
| In operating a motor vehicle | 0.89 | 1.41 | 3.42 | 2.28 | 0.29 | 0.81 |
| In my relationships with friends | 2.18 | 2.04 | 4.40 | 1.87 | 0.96 | 1.31 |
| In my marital, or partner, or dating relationships | 2.61 | 2.41 | 4.44 | 2.08 | 0.98 | 1.56 |
| In my educational classes (e.g., attendance) | 4.63 | 2.67 | 5.72 | 1.96 | 1.45 | 1.67 |
| In my performance on educational tests/assignments | 4.95 | 2.74 | 6.23 | 1.69 | 1.76 | 1.61 |
| In controlling my behavior at work, home, or school | 3.29 | 2.17 | 5.65 | 1.69 | 0.55 | 1.10 |
| In my decision making at work, home, or school | 3.42 | 2.06 | 5.09 | 1.68 | 1.16 | 1.42 |
| In maintaining hygiene (dressing, showing) | 1.58 | 1.64 | 3.14 | 2.14 | 0.49 | 1.23 |
| In self-care (e.g., sleeping, eating) | 3.29 | 2.38 | 4.04 | 2.30 | 1.33 | 1.76 |
| In social activities | 2.34 | 1.92 | 4.63 | 1.93 | 0.85 | 1.16 |
| In community-based activities (e.g., church, clubs, organizations) | 2.05 | 1.83 | 4.91 | 1.94 | 0.85 | 1.45 |
| In maintaining my health (e.g., nutrition, exercise) | 3.34 | 2.47 | 3.91 | 2.19 | 1.69 | 1.90 |
| In time management | 5.71 | 2.17 | 6.35 | 1.65 | 1.96 | 1.92 |
| In meeting deadlines | 4.24 | 2.55 | 6.18 | 1.77 | 1.20 | 1.68 |
| With controlling my anger | 3.16 | 2.52 | 4.49 | 2.08 | 0.75 | 1.46 |
| With my memory for daily activities | 4.34 | 2.18 | 5.91 | 1.73 | 0.98 | 1.39 |

Table 19

*Means and Standard Deviations of the Original Symptom Validity Index 7*

| | ADHD ($n$ = 38) | | Malingering ($n$ = 57) | | Control ($n$ = 55) | |
|---|---|---|---|---|---|---|
| Symptom Validity Item | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| SV-Item #1 | 1.87 | 1.82 | 4.44 | 2.39 | .36 | .65 |
| SV-Item #2 | 3.16 | 1.90 | 4.37 | 2.34 | .31 | .74 |
| SV-Item #3 | 1.92 | 1.95 | 4.44 | 1.95 | .84 | 1.30 |
| SV-Item #4 | 2.24 | 2.41 | 4.09 | 1.92 | .38 | .91 |
| SV-Item #5 | 2.29 | 2.21 | 4.12 | 2.10 | .55 | .29 |
| SV-Item #6 | 1.97 | 1.99 | 3.98 | 2.19 | .40 | .71 |
| SV-Item #7 | 1.84 | 2.07 | 4.19 | 2.18 | .44 | 1.09 |

Table 20

*Means and Standard Deviations of the Revised Symptom Validity Index 6*

| | ADHD ($n$ = 38) | | Malingering ($n$ = 57) | | Control ($n$ = 55) | |
|---|---|---|---|---|---|---|
| Symptom Validity Item | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| SV-Item #2 | 3.16 | 1.90 | 4.37 | 2.34 | .31 | .74 |
| SV-Item #3 | 1.92 | 1.95 | 4.44 | 1.95 | .84 | 1.30 |
| SV-Item #4 | 2.24 | 2.41 | 4.09 | 1.92 | .38 | .91 |
| SV-Item #1 | 1.87 | 1.82 | 4.44 | 2.39 | .36 | .65 |
| *SV-Item #8* | 2.32 | 1.69 | 5.11 | 1.89 | .49 | .92 |
| *SV-Item #9* | 2.68 | 1.99 | 5.11 | 1.82 | .73 | 1.06 |

*Note:* New items are italicized.

Table 21

*Correlation Matrix of Primary Measures*

| | I-index | HI-index | Total Index | FI-Index | SV-Index 7 | SV-Index 6 | CAT-A Item | CAT-A Total | WMT IR | WMT DR | WMT CNS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HI-index | .90 | - | - | - | - | - | - | - | - | - | - |
| Total Symptom Index | .98 | .97 | - | - | - | - | - | - | - | - | - |
| FI-Index | .87 | .86 | .89 | - | - | - | - | - | - | - | - |
| SV-Index 7 | .81 | .88 | .87 | .85 | - | - | - | - | - | - | - |
| SV-Index 6 | .85 | .90 | .89 | .88 | .98 | - | - | - | - | - | - |
| CAT-A Item | .58 | .64 | .62 | .60 | .64 | .64 | - | - | - | - | - |
| CAT-A Total | .81 | .88 | .87 | .81 | .81 | .83 | .71 | - | - | - | - |
| WMT IR | -.29 | -.33 | -.31 | -.36 | -.32 | -.33 | -.20 | -.33 | - | - | - |
| WMT DR | -.38 | -.40 | -.40 | -.42 | -.40 | -.40 | -.20 | -.38 | .65 | - | - |
| WMT CNS | -.38 | -.41 | -.41 | -.43 | -.40 | -.40 | -.18 | -.39 | .71 | .95 | - |
| WMT MC | -.40 | -.41 | -.42 | -.41 | -.38 | -.39 | -.20 | -.39 | .68 | .90 | .90 |

Note: All correlations are significant $p < .001$

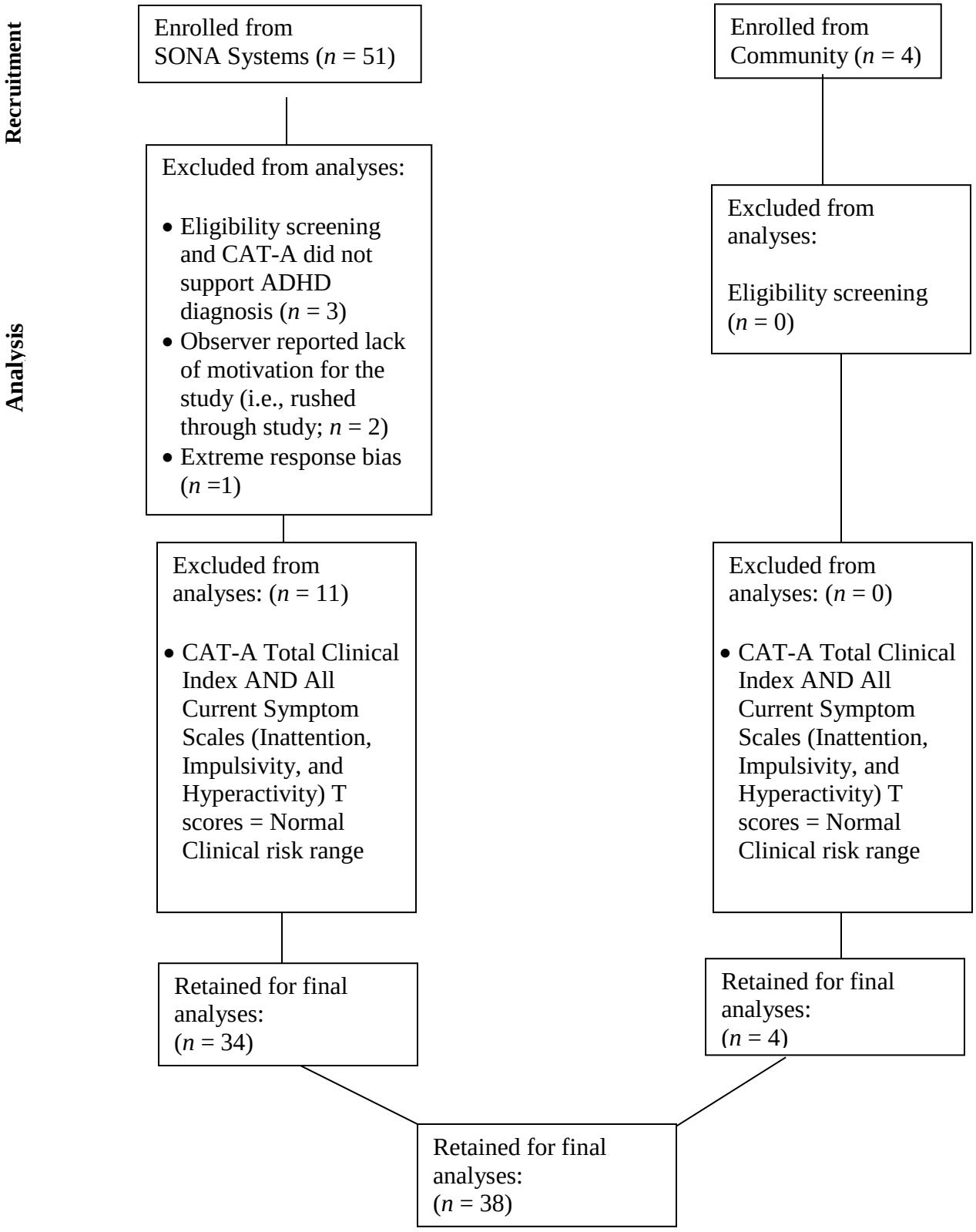


*Figure 1.* CONSORT Diagram ADHD Participants

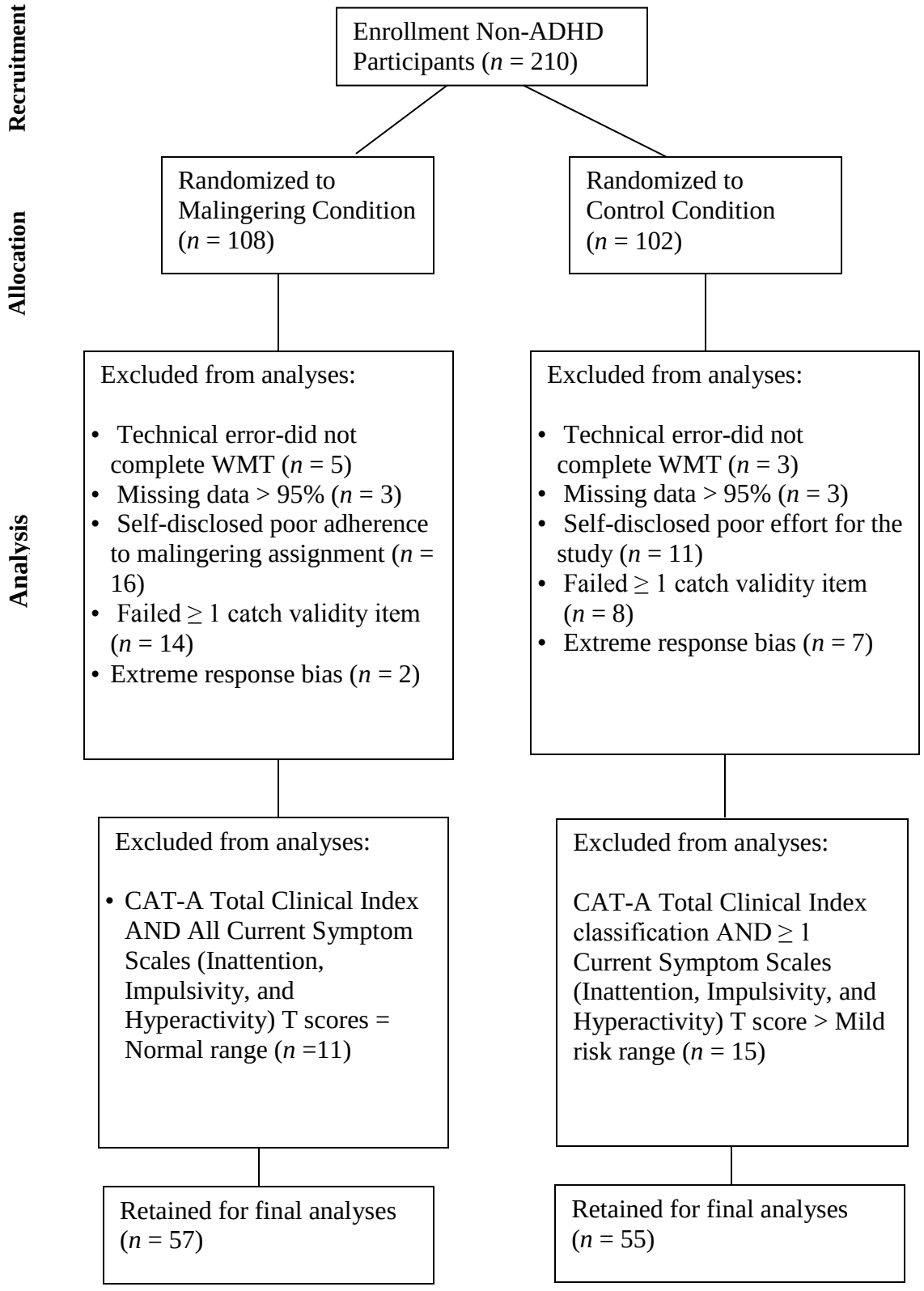


*Figure 2.* CONSORT Diagram Non-ADHD Participan

# Appendix A

## Diagnostic Criteria for ADHD in the *Diagnostic and Statistical Manual for Mental Disorders-Fifth Edition*

***Diagnostic Criteria***

A. A persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development, as characterized by (1) and/or (2):

1. **Inattention:** Six (or more) of the following symptoms have persisted for at least 6 months to a degree that is inconsistent with developmental level and that negatively impacts directly on social and academic/occupational activities:
    - **Note:** The symptoms are not solely a manifestation of oppositional behavior, defiance, hostility, or failure to understand tasks or instructions. For older adolescents and adults (age 17 and older), at least five symptoms are required.
    a. Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities (e.g., overlooks or misses details, work is inaccurate).
    b. Often has difficulty sustaining attention in tasks or play activities (e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading).
    c. Often does not seem to listen when spoken to directly (e.g., mind seems elsewhere, even in the absence of any obvious distraction).
    d. Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., starts tasks but quickly loses focus and is easily sidetracked).
    e. Often has difficulty organizing tasks and activities (e.g., difficulty managing sequential tasks; difficulty keeping materials and belongings in order; messy, disorganized work; has poor time management; fails to meet deadlines).
    f. Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework; for older adolescents and adults, preparing reports, completing forms, reviewing lengthy papers).
    g. Often loses things necessary for tasks or activities (e.g., school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).
    h. Is often easily distracted by extraneous stimuli (for older adolescents and adults, may include unrelated thoughts).
    i. Is often forgetful in daily activities (e.g., doing chores, running errands; for older adolescents and adults, returning calls, paying bills, keeping appointments).
2. **Hyperactivity and impulsivity:** Six (or more) of the following symptoms have persisted for at least 6 months to a degree that is inconsistent with developmental level and that negatively impacts directly on social and academic/occupational activities:
    - **Note:** The symptoms are not solely a manifestation of oppositional behavior, defiance, hostility, or a failure to understand tasks or instructions. For older adolescents and adults (age 17 and older), at least five symptoms are required.
    a. Often fidgets with or taps hands or feet or squirms in seat.
    b. Often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplace, or in other situations that require remaining in place).
    c. Often runs about or climbs in situations where it is inappropriate. (**Note:** In adolescents or adults, may be limited to feeling restless.)
    d. Often unable to play or engage in leisure activities quietly.
    e. Is often "on the go," acting as if "driven by a motor" (e.g., is unable to be or uncomfortable being still for extended time, as in restaurants, meetings; may be experienced by others as being restless or difficult to keep up with).
    f. Often talks excessively.
    g. Often blurts out an answer before a question has been completed (e.g., completes people's sentences; cannot wait for turn in conversation).
    h. Often has difficulty waiting his or her turn (e.g., while waiting in line).

i. Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission; for adolescents and adults, may intrude into or take over what others are doing).

B. Several inattentive or hyperactive-impulsive symptoms were present prior to age 12 years.
C. Several inattentive or hyperactive-impulsive symptoms are present in two or more settings (e.g., at home, school, or work; with friends or relatives; in other activities).
D. There is clear evidence that the symptoms interfere with, or reduce the quality of, social, academic, or occupational functioning.
E. The symptoms do not occur exclusively during the course of schizophrenia or another psychotic disorder and are not better explained by another mental disorder (e.g., mood disorder, anxiety disorder, dissociative disorder, personality disorder, substance intoxication or withdrawal).

*Specify* whether:

- **314.01 (F90.2) Combined presentation:** If both Criterion A1 (inattention) and Criterion A2 (hyperactivity-impulsivity) are met for the past 6 months.
- **314.00 (F90.0) Predominantly inattentive presentation:** If Criterion A1 (inattention) is met but Criterion A2 (hyperactivity-impulsivity) is not met for the past 6 months.
- **314.01 (F90.1) Predominantly hyperactive/impulsive presentation:** If Criterion A2 (hyperactivity-impulsivity) is met and Criterion A1 (inattention) is not met for the past 6 months.

*Specify* if:

- **In partial remission:** When full criteria were previously met, fewer than the full criteria have been met for the past 6 months, and the symptoms still result in impairment in social, academic, or occupational functioning.

*Specify* current severity:

- **Mild:** Few, if any, symptoms in excess of those required to make the diagnosis are present, and symptoms result in no more than minor impairments in social or occupational functioning.
- **Moderate:** Symptoms or functional impairment between "mild" and "severe" are present.
- **Severe:** Many symptoms in excess of those required to make the diagnosis, or several symptoms that are particularly severe, are present, or the symptoms result in marked impairment in social or occupational functioning.

Appendix B

Study Flyer

# RECRUITING
# ADULTS WITH ADHD
# BETWEEN THE AGES OF 18-26

What is the study?

- Research on how to accurately detect ADHD
- Complete rating scales and a brief memory test
- Takes 45-60 minutes
- All information is kept confidential
-

What is the compensation?

- Can earn up to $40 cash
- May be entered into a $100 Gift Card raffle drawing

Where do I complete the study?

- Private testing room

For more information and
to complete the eligibility screening please contact:
Heather Potts

Appendix C

ADHD Screening Form

**ADHD Screening Form**

Date/Time of Screening/Study _____
Completed by _____

☐ Provide Overview of Study Prior to Screening (*script below)*
*"Thank you for your interest in participating in this study. We are interested in learning more about how to accurately diagnose ADHD in college students.*
*For this part of the study, we are recruiting individuals between the ages of 18-26 and those who have a diagnosis of ADHD. In order to verify your eligibility to participate, we have some brief screening questions to ask you. But first, let me tell you about the study. If you are eligible and decide to participate, you will be asked to set up a convenient time to complete the study in person. As part of this study, you will be asked to complete a brief demographic survey to collect some background information about you, including questions about diagnoses, accommodations, and medication usage. Next, you will complete a computerized memory test and three ADHD symptom and impairment rating measures. We ask that you complete each item honestly as they pertain to you. The final measure will be a brief exit survey about your experience, effort during the study, along with previous knowledge about ADHD. The entire study will take approximately 60 minutes of your time*

*All information will be kept confidential. This means that your name will not appear anywhere and your specific answers will not be linked to your name in any way. Your Screening ID/name will only be connected to reimbursement. In addition, all of your information will be used for research purposes only. We also believe that this study should involve minimal risk to you. Taking part in this research study is optional and your decision and you have the right to both participate, and the option to opt-out at any time.*

*In compensation for your time, participants who return the completed packet will receive up to 1 SONA credit hour, and may be entered into a raffle drawing for a $100 Visa gift card. The researchers will notify the winner of the raffle directly after the completion of the study.*

*That is an overview of the study. Would you like to continue with the phone screening to determine if you are eligible?* ☐Yes ☐ No

***If NO***
*"Thank you for your time."* Politely discontinue screening.

**NEXT PAGE IF YES**

***If YES***

*"Thank you for your continued interest"*
*Assign a Screening ID# and Proceed with Questions below*

Screening ID#_____

| | |
|---|---|
| ***Assign Screening ID #***<br>*Write Name of Participant and Screening ID number on Excel file* | _____ |
| 1. How did you hear about the study? (*write in)* | _____ |
| 2. Are you between the ages of 18-26? | ☐ Yes ☐ No |
| 3. Do you have a diagnosis of ADHD? | ☐ Yes ☐ No |
| 4. Were you diagnosed by a professional?<br><br>If yes, what type of professional? (*write in)*<br><br>*Additional Probes: Ask whether they completed testing (do you remember how you were diagnosed? Did you complete any tasks, or fill out any rating scales?)*<br><br>*Provide prompts if needed "was it a doctor, psychologist, school professional?"* | ☐ Yes ☐ No<br><br>_____ |
| 5. At what age were you diagnosed? (*write in)*<br><br>*Additional Probes:*<br>*Ask about grade level if they are unsure. At least get an estimate, such as elementary, middle, or college* | _____ |
| 6. Did you experience ADHD symptoms prior to the age of 12? | ☐ Yes ☐ No |
| 7. Do you still experience symptoms of ADHD?<br><br>*Additional Probes:*<br>*Ask about specific symptoms, "What types of symptoms did you experience currently" "Are these symptoms the same or different in comparison to childhood?"* | ☐ Yes ☐ No |

| **Impairment** | |
|---|---|
| 8. In the past 6 months, have your symptoms impacted you in your everyday life? In other words, have your symptoms caused you any impairment or difficulty?<br>• *DSM-V Definition of Impairment- "Evidence that symptoms interfere with, or reduce the quality of social, academic, or occupational functioning"*<br>• *Document evidence of impairment in Notes*<br>• *Check box if there is evidence of impairment*<br>• *Provide additional clarification and examples if needed* | ☐ Yes ☐ No<br>*Check if individual demonstrates impairment in at least ONE area* |
| <u>Social Functioning:</u><br>*Do your ADHD symptoms sometimes cause difficulty with your social life or conversations with others?* | ☐ Yes, Evidence of Impairment |
| <u>Academic Functioning:</u><br>*Do your ADHD symptoms impact your functioning in the classroom, or completing school work at home?"* | ☐ Yes, Evidence of Impairment |
| <u>Occupational Functioning:</u><br>*Do you have a job? Did symptoms experience any difficulties maintaining a job?* | ☐ Yes, Evidence of Impairment |
| <u>Home Life</u><br>*Do your symptoms impact your ability to manage every day life, such as managing money, organizing, planning, etc.?* | |
| Notes about symptoms and impairment | |

☐ Not Eligible

*"Thank you for your interest and participating in this phone screening. At this time, you do not meet our eligibility criteria to participate in this study. Again, I appreciate your interest and should you have any questions, please do not hesitate to let me know."*

☐ Eligible if:

- "Yes" is checked for #2, #3, #6, and #8
- Diagnosis was made by a qualified mental health provider, psychologist, counselor

**NEXT STEPS:**

- Complete remainder of the study using the ADHD SONA study protocol

Appendix D

Demographic Questionnaire

Age: _____
Gender: ____M ____F ____Other
SU ID or Screening ID:_____
Current GPA: _____

Year in School *(Please check)*
____Freshman ___Sophomore ___Junior __Senior ___5+ years ___Graduate ___Not a student

Ethnicity: (*Please check)*
____American Indian or Alaska Native ____ Asian ____ Black or African American
____ Hispanic or Latino ____ Native Hawaiian or Other Pacific Islander ____ White

Primary language:
____English _____Other_____(please write/type first language)

Are you currently experiencing any difficulties related to school?
_____No _____Yes _____N/A
If Yes, please explain:_____ _____

Please check any disorder with which you have been **diagnosed**
____ADHD/ADD _____Anxiety Disorder
____Learning Disability _____ Traumatic Brain Injury
____Vision Impairment (**uncorrected with glasses*) _____ Autism
____Hearing Impairment (**uncorrected with hearing aids*) _____ Other:_____
____Depression _____ None

Are you currently taking any medications?
_____No _____Yes
If so, what is (are) the medication(s) treating? _____

Have you ever received academic and/or testing accommodations in high school or college (e.g., extended time, tests in a private room)
_____No _____Yes

Do you regularly take stimulant medication (e.g., Ritalin, Adderall)?
_____No _____Yes
If Yes, did you take stimulant medication within 12 hours of participating in this study?
_____No _____Yes

Have you ever tried stimulant medication before (with or without a prescription)?
_____No _____Yes

Do you think you have an ADHD diagnosis?
_____No _____Yes

Have you ever had a prior evaluation to determine whether you have ADHD?
_____No _____Yes

Appendix E

Exit Survey-ADHD and Control Groups

I believe that I put forth my best effort on the self-report measure

(Please Circle) No Yes

If No, Please Explain:_____

I completed the research materials (e.g. self-report measure) honestly and accurately

(Please Circle) No Yes

If No, Please Explain:_____

*Prior to this study,* were you aware of the symptoms of ADHD

(Please Circle) No Yes

*Prior to this study*, I would rank my knowledge of ADHD on a scale of 0-8 as:

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| No Knowledge | | Little | | Good | | Very Good | | Superior Knowledge |

*Prior to this study*, I learned about ADHD from:

(Check all that apply)

____I have ADHD ____TV Advertisements

____Family member has ADHD ____Brochures/Pamphlets

____Friend has ADHD ____News reports

____Research into ADHD (Journal Articles) ____Online websites

____I had no knowledge of ADHD before this study

Appendix F

Exit Survey-Malingering Group

I believe that I attempted to fake ADHD to the best of my ability
No Yes
If No, Please explain

I believe that I was successful in faking ADHD
(Please Circle) No Yes
If No, Please Explain:_____

I completed the research materials (e.g. self-report measure) as if I was faking ADHD
(Please Circle) No Yes
If No, Please Explain:_____

*Prior to this study*, were you aware of the symptoms of ADHD
(Please Circle) No Yes

*Prior to this study*, I would rank my knowledge of ADHD on a scale of 0-8 as:

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| No | | Little | | Good | | Very Good | | Superior |

*Prior to this study*, I knew about ADHD from:
(Check all that apply)
____I have ADHD ____TV Advertisements
____Family member has ADHD ____Brochures/Pamphlets
____Friend has ADHD ____News reports
____Research into ADHD (Journal Articles) ____Online websites
____I had no knowledge of ADHD before this study

I feel like the ADHD information packet at the beginning of the study was necessary for me to read in order to successfully fake the symptoms of ADHD
No Yes

I used the following strategies while taking this ADHD self-report measure
(Check all that apply)
____Selected items that best matched the DSM-V Criteria (provided in handout)
____ Selected items that best matched my previous knowledge of ADHD
____ Selected items that best matched a person I know with ADHD
____ Impulsive with response selection
____ Did not read instructions fully
____ Completed tasks slowly
____ Skipped items
____Re-read items
____ Selected items about Inattention
____ Selected items about Hyperactivity
____ Letting mind wander or "zoning out"
____ Other:_____

References

Aldridge, A. P., Kroutil, L. A., Cowell, A. J., Reeves, D. B., & Van Brunt, D. L. (2011). Medication costs to private insurers of diversion of medications for attention-deficit hyperactivity disorder. *Pharmacoeconomics*, *29*, 621–635.

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

Americans with Disabilities Act Amendments Act of 2008, Pub. L. No. 110–325, 3406 (2008).

Barkley, R., Murphy, K., & Fischer, M. (2008). *ADHD in adults: What the science says*. New York, NY: Guilford Press.

Benson, K., Flory, K., Humphreys, K. L., & Lee, S. S. (2015). Misuse of stimulant medication among college students: A comprehensive review and meta-analysis. *Clinical Child and Family Psychology Review*, *18*, 50–76. https://doi.org/10.1007/s10567-014-0177-z

Bird, H. R., Canino, G., Rubio-Stipec, M., Gould, M. S., Ribera, J., Sesman, M., … Moscoso, M. (1988). Estimates of the prevalence of childhood maladjustment in a community survey in Puerto Rico: The use of combined measures. *Archives of General Psychiatry*, *45*, 1120–1126. https://doi.org/10.1001/archpsyc.1988.01800360068010

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. https://doi.org/10.1177/1087054708329927

Bracken, B., & Boatwright, B. (2005). *CAT-A, Clinical Assessment of Attention Deficit-Adult: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Bracken, B., & Howell, K. (2004). *Clinical Assessment of Depression: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Brennan, A. M., & Gouvier, W. D. (2006). Are we honestly studying malingering? A profile and comparison of simulated and suspected malingerers. *Applied Neuropsychology*, *13*, 1–11. https://doi.org/10.1207/s15324826an1301_1

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., … Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*, 419–426. https://doi.org/10.1016/j.acn.2005.02.002

Bussing, R., M. D. .. M. S. H. ., Zima, B. T., M. D. .. M. P. ., Mason, D. M., B. ., Porter, P. C., B. ., & Garvan, C. W., Ph. . (2011). Receiving treatment for Attention-Deficit Hyperactivity Disorder: Do the perspectives of adolescents matter? *Journal of Adolescent Health*, *49*(1), 7–14. https://doi.org/10.1016/j.jadohealth.2010.08.014

Centers for Disease Control and Prevention. (2014). Attention Deficit/Hyperactivity Disorder: Data and Statistics.

Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., … Ord, J. S. (2015). Official position of the American Academy of Clinical Neuropsychology: Social Security Administration policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist*, *29*, 723–740. https://doi.org/10.1080/13854046.2015.1099738

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J: L. Erlbaum Associates.

Conners, C. K., Erhardt, D., & Sparrow, E. P. (1998). *Conners' Adult ADHD Rating Scales: Self-Report*. North Tonawanda, NY: Multi-Health Systems, Inc.

Cook, C. M., Bolinger, E., & Suhr, J. A. (2016). Further validation of the Conner's Adult Attention Deficit/Hyperactivity Rating Scale Infrequency Index (CII) for detection of non-credible report of Attention Deficit/Hyperactivity Disorder symptoms. *Archives of Clinical Neuropsychology*, *31*, 358–364. https://doi.org/10.1093/arclin/acw015

Davidson, M. A. (2008). ADHD in adults: A review of the literature. *Journal of Attention Disorders*, *11*, 628–641. https://doi.org/10.1177/1087054707310878

DuPaul, G. J., Reid, R., Anastopoulos, A. D., & Power, T. J. (2014). Assessing ADHD symptomatic behaviors and functional impairment in school settings: Impact of student and teacher characteristics. *School Psychology Quarterly*, *29*, 409–421. https://doi.org/10.1037/spq0000095

Dvorsky, M. R., Langberg, J. M., Molitor, S. J., & Bourchtein, E. (2016). Clinical utility and predictive validity of parent and college student symptom ratings in predicting an ADHD diagnosis. *Journal of Clinical Psychology*, *72*, 401–418. https://doi.org/10.1002/jclp.22268

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

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

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

Gathje, R. A., Lewandowski, L. J., & Gordon, M. (2008). The role of impairment in the diagnosis of ADHD. *Journal of Attention Disorders*, *11*, 529–537.

George, D., & Mallery, P. (2009). *SPSS for Windows step by step: a simple guide and reference, 16.0 update* (9th ed.). Boston, MA: Allyn and Bacon.

Gordon, M., Antshel, K., Faraone, S., Barkley, R., Lewandowski, L., Hudziak, J. J., … Cunningham, C. (2006). Symptoms versus impairment: The case for respecting DSM-IV's criterion D. *Journal of Attention Disorders*, *9*, 465–475. https://doi.org/10.1177/1087054705283881

Green, P. (2003). *The Word Memory Test*. Edmonton/Seattle: Green's Publishing.

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the word memory test and medical symptom validity test in groups with severe verbal memory impairment. *Applied Neuropsychology*, *18*, 86–94. https://doi.org/10.1080/09084282.2010.523389

Harrison, A. G. (2004). An Investigation of reported symptoms of ADHD in a university population. *The ADHD Report*, *12*, 8–11. https://doi.org/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. https://doi.org/10.1007/s12207-017-9291-9

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*, 265–283. https://doi.org/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*, 135–143. https://doi.org/10.1080/09084281003715642

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference, P. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *Clinical Neuropsychologist*, *23*, 1093–1129. https://doi.org/10.1080/13854040903155063

Hirsch, O., & Christiansen, H. (2015). Faking ADHD? Symptom validity testing and Its relation to self-reported, observer-reported symptoms, and neuropsychological measures of atention in adults With ADHD. *J Atten Disord*. https://doi.org/10.1177/1087054715596577

Ilieva, I. P., Hook, C. J., & Farah, M. J. (2015). Prescription stimulants' efects on healthy inhibitory control, working memory, and episodic memory: A meta-analysis. *Journal of Cognitive Neuroscience*, *27*(6), 1069–1089. https://doi.org/10.1162/jocn_a_00776

Iverson, G. L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology*, *13*, 77–90. https://doi.org/10.1207/s15324826an1302_3

Jasinski, L. J., Harp, J. P., Berry, D. T. R., 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. https://doi.org/10.1080/13854046.2011.630024

Jasinski, L. J., & Ranseen, J. D. (2011). Malingered ADHD evaluations: A further complication for accommodations reviews. *The Bar Examiner*, *80*, 6–16.

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*, 716–723. https://doi.org/10.1176/appi.ajp.163.4.716

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*, 625–631. https://doi.org/10.1017/S1355617712000240

Lasky, A. K., Weisner, T. S., Jensen, P. S., Hinshaw, S. P., Hechtman, L., Arnold, L. E., … Swanson, J. M. (2016). ADHD in context: Young adults' reports of the impact of occupational environment on the manifestation of ADHD. *Social Science and Medicine*, *161*, 160–168. https://doi.org/10.1016/j.socscimed.2016.06.003

Lebowitz, M. S. (2016). Stigmatization of ADHD: A developmental review. *Journal of Attention Disorders*, *20*(3), 199–205. https://doi.org/10.1177/1087054712475211

Lewandowski, L. J., Lambert, T. L., Lovett, B. J., Panahon, C. J., & Sytsma, M. R. (2014). College students' preferences for test accommodations. *Canadian Journal of School Psychology*, *29*, 116–126. https://doi.org/10.1177/0829573514522116

Lewandowski, L. J., Lovett, B. J., Codding, R. S., & Gordon, M. (2008). Symptoms of ADHD and academic concerns in college students with and without ADHD diagnoses. *Journal of Attention Disorders*, *12*, 156–161. https://doi.org/10.1177/1087054707310882

Marshall, P., Hoelzle, J. B., Heyerdahl, D., & Nelson, N. W. (2016). The impact of failing to identify suspect effort in patients undergoing Adult Attention-Deficit/Hyperactivity Disorder (ADHD) assessment. *Psychological Assessment*, *28*(10), 1290–1302. https://doi.org/10.1037/pas0000247

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*, 1204–1237. https://doi.org/10.1080/13854046.2010.514290

McLeod, J. D. (2007). Public knowledge, beliefs, and treatment preferences concerning attention-deficit hyperactivity disorder. *Psychiatric Services*, *58*, 626–631. https://doi.org/10.1176/appi.ps.58.5.626

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

Nelson, J. M., Whipple, B., Lindstrom, W., & Foels, P. A. (2014). How is ADHD assessed and documented? Examination of psychological reports submitted to determine eligibility for postsecondary disability. *Journal of Attention Disorders*. https://doi.org/10.1177/1087054714561860

Owens, J. S., Goldfine, M. E., Evangelista, N. M., Hoza, B., & Kaiser, N. M. (2007). A critical review of self-perceptions and the positive illusory bias in children with ADHD. *Clinical Child and Family Psychology Review*, *10*, 335–351. https://doi.org/10.1007/s10567-007-0027-3

Park, T. M., & Haning, W. F. (2016). Stimulant use disorders. *Child and Adolescent Psychiatric Clinics of North America, 25*, 461–471. https://doi.org/10.1016/j.chc.2016.02.004

Pescosolido, B. A., Jensen, P., Martin, J., Perry, B., Olafsdottir, S., & Fettes, D. (2008). Public knowledge and assessment of child mental health problems: Findings From the National Stigma Study-Children. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(3), 339–349. https://doi.org/10.1097/CHI.0b013e318160e3a0

Potts, H. E. (2016). *The construction of a multidimensional ADHD self-report scale: A pilot study (unpublished master's thesis)*. Syracuse University, Syracuse, New York.

Prevatt, F., Proctor, B., Best, L., Baker, L., Van Walker, J., & Taylor, N. W. (2012). The positive illusory bias: Does it explain self-evaluations in college students with ADHD? *Journal of Attention Disorders, 16*, 235–243. https://doi.org/10.1177/1087054710392538

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

Raue, K., & Lewis, L. (2011). *Students with disabilities at degree granting postsecondary institutions*. US Department of Education National Center for Education Statistics.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior Assessment System for Children, Third Edition*. Toronto, Canada: Pearson.

Rogers, R. (2008). *Clinical assessment of malingering and deception (3rd ed.)*. New York, NY, US: Guilford Press.

Rogers, R., & Gillard, N. D. (2011). Research methods for the assessment of malingering. In B. Rosenfeld & S. D. Penrod (Eds.), *Research methods in forensic psychology*. (pp. 174–188). Hoboken, NJ, US: John Wiley & Sons Inc.

Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A Comparison of Forensic and Nonforensic Malingerers: A Prototypical Analysis of Explanatory Models. *Law and Human Behavior*, *22*(4), 353–367. https://doi.org/10.1023/A:1025714808591

Rogers, R., Sewell, K. W., Morey, L. C., & Ulstad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment, 67*, 629.

Sagar, S., Miller, C. J., & Erdodi, L. A. (2017). Detecting feigned attention-deficit/hyperactivity disorder (ADHD): Current methods and future directions. *Psychological Injury and Law*, *10*, 105–113. https://doi.org/10.1007/s12207-017-9286-6

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 28–38. https://doi.org/10.1097/00004583-200001000-00014

Slick, D. J., Hopp, G., Strauss, E., & Thompson, G. B. (1996). *VSVT: Victoria Symptom Validity Test, version 1.0, professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)*, *13*, 545–561. https://doi.org/10.1076/1385-4046(199911)13:04;1-y;ft545

Sollman, M. J., & Berry, D. T. (2011). Detection of inadequate effort on neuropsychological testing: a meta-analytic update and extension. *Archives of Clinical Neuropsychology*, *26*, 774–789. https://doi.org/10.1093/arclin/acr066

Sonuga-Barke, E. J. (2005). Causal models of attention-deficit/hyperactivity disorder: from common simple deficits to multiple developmental pathways. *Biological Psychiatry*, *57*, 1231–1238. https://doi.org/10.1016/j.biopsych.2004.09.008

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary* (3rd Edition). Oxford, NY: Oxford University Press.

Suhr, J. A., Buelow, M., & Riddle, T. (2011). Development of an Infrequency Index for the CAARS. *Journal of Psychoeducational Assessment*, *29*, 160–170. https://doi.org/10.1177/0734282910380190

Suhr, J. A., Cook, C., & Morgan, B. (2017). Assessing functional impairment in ADHD: Concerns for validity of self-report. *Psychological Injury and Law*, *10*, 151–160. https://doi.org/10.1007/s12207-017-9292-8

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

Suhr, J. A., Sullivan, B. K., & Rodriguez, J. L. (2011). The relationship of noncredible performance to Continuous Performance Test scores in adults referred for attention-deficit/hyperactivity disorder evaluation. *Archives of Clinical Neuropsychology*, *26*, 1–7. https://doi.org/10.1093/arclin/acq094

Suhr, J. A., & Wei, C. (2013). Symptoms as an excuse: Attention deficit/hyperactivity disorder symptom reporting as an excuse for cognitive test performance in the context of evaluative threat. *Journal of Social and Clinical Psychology*, *32*, 753–769. https://doi.org/10.1521/jscp.2013.32.7.753

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. https://doi.org/10.1080/09084280701509083

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, *99*(1), 100–117. https://doi.org/10.1037/0033-2909.99.1.100

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*, S123–S134. https://doi.org/10.1007/s00702-014-1274-3

Tucha, L., Sontag, T. A., Walitza, S., & Lange, K. W. (2009). Detection of malingered attention deficit hyperactivity disorder. *ADHD Attention Deficit Hyperactivity Disorders*, *1*, 47–53. https://doi.org/10.1007/s12402-009-0007-1

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *Clinical Neuropsychologist*, *23*, 297–313. https://doi.org/10.1080/13854040802232682

Wakefield, J. C. (2010). Misdiagnosing normality: Psychiatry's failure to address the problem of false positive diagnoses of mental disorders in a changing professional environment. *Journal of Mental Health*, *19*, 337–351.

Weyandt, L. L., & DuPaul, G. (2006). ADHD in college students. *Journal of Attention Disorders, 10,* 9–19. https://doi.org/10.1177/1087054705286061

**Heather Potts**
**Vita**

**Contact**
pottshe@gmail.com

**Education**

Syracuse University — Syracuse, NY
**Doctorate of Philosophy** — August 2018
School Psychology, APA, CAEP Accredited and NASP Approved
Dissertation: The Detection of ADHD and Malingering in Young Adults
Advisor: Lawrence Lewandowski, Ph.D.

Syracuse University — Syracuse, NY
**Master of Science** — August 2016
Thesis: The Construction of a Multidimensional ADHD Rating Scale
Advisor: Lawrence Lewandowski, Ph.D.

The George Washington University — Washington, DC
**Master of Arts in Education & Human Development** — August 2008
Rehabilitation Counseling

James Madison University — Harrisonburg, VA
**Bachelor of Arts** — May 2006
Cum Laude
Major: Psychology; Minor: Modern Foreign Languages-Spanish