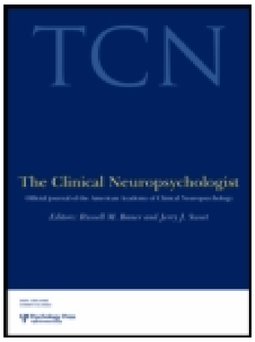

Routledge
Taylor & Francis Group

The Clinical Neuropsychologist

ISSN: 1385-4046 (Print) 1744-4144 (Online) Journal homepage: https://www.tandfonline.com/loi/ntcn20

Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder

Paul Marshall , Ryan Schroeder , Jeffrey O'Brien , Rebecca Fischer , Adam Ries , Brita Blesi & Jessica Barker

The Clinical Neuropsychologist, 24: 1204–1237, 2010
http://www.psypress.com/tcn
ISSN: 1385-4046 print/1744-4144 online
DOI: 10.1080/13854046.2010.514290

Psychology Press
Taylor & Francis Group

EFFECTIVENESS OF SYMPTOM VALIDITY MEASURES IN IDENTIFYING COGNITIVE AND BEHAVIORAL SYMPTOM EXAGGERATION IN ADULT ATTENTION DEFICIT HYPERACTIVITY DISORDER

Paul Marshall[1], Ryan Schroeder[1,2], Jeffrey O'Brien[1], Rebecca Fischer[1], Adam Ries[1], Brita Blesi[1], and Jessica Barker[1]
[1]Department of Psychiatry, Hennepin County Medical Center, Minneapolis, MN and [2]University of Kansas School of Medicine, Wichita, KS, USA

This study examines the effectiveness of symptom validity measures to detect suspect effort in cognitive testing and invalid completion of ADHD behavior rating scales in 268 adults referred for ADHD assessment. Patients were diagnosed with ADHD based on cognitive testing, behavior rating scales, and clinical interview. Suspect effort was diagnosed by at least two of the following: failure on embedded and free-standing SVT measures, a score > 2 SD below the ADD population average on tests, failure on an ADHD behavior rating scale validity scale, or a major discrepancy between reported and observed ADHD behaviors. A total of 22% of patients engaged in symptom exaggeration. The Word Memory test immediate recall and consistency score (both 64%), TOVA omission errors (63%) and reaction time variability (54%), CAT-A infrequency scale (58%), and b Test (47%) had good sensitivity as well as at least 90% specificity. Clearly, such measures should be used to help avoid making false positive diagnoses of ADHD.

Keywords: ADHD; Adults; Effort testing; Symptom exaggeration; Symptom validity.

INTRODUCTION

Extensive research has clearly established that malingering—the intentional exaggeration or fabrication of symptoms including cognitive deficits—is a relatively common occurrence in neuropsychological evaluations. Approximately 40% of patients seen for medico-legal evaluations will malinger during neuropsychological testing (Larrabee, 2003). Boone et al. estimate that approximately 30% of patients in settings that receive referrals from a mixture of general clinical and medico-legal sources will malinger during neuropsychological assessment (Boone, Lu, & Herzberg, 2002b). It is also estimated that 15% of individuals might produce invalid performances on tests of cognitive functioning, even in general clinical settings where patients seeking compensation are considered to be rarer (Rogers, Harrell, & Liff, 1993). For this reason, the clinical practice guidelines promulgated by both the American Academy of Clinical Neuropsychology (AACN) and the

Address correspondence to: Paul Marshall, Hennepin County Medical Center, Department of Psychiatry, Room S110, 914 South 8th Street, Minneapolis, MN 55404, USA. E-mail: paul.marshall@hfanet.org
Accepted for publication: July 12, 2010. First published online: September 13, 2010

National Academy of Neuropsychology state that the identification of suspect effort in neuropsychological testing should be an integral component of a valid assessment of a patient's cognitive function (AACN, 2007; Bush et al., 2005).

There is ample reason to be concerned that a significant number of patients referred for assessment of ADHD might exaggerate or fabricate not only cognitive deficits in testing but also behavioral problems on self-report measures in an effort to receive that diagnosis. In general, individuals might be tempted to feign symptoms of ADHD when they are facing the possibility of failing in school or in their job. Students in particular might seek the ADHD diagnosis when confronted with having difficulty with the significantly greater demands of college or graduate school course work (Frazier, Frazier, Busch, Kerwood, & Demaree, 2008).

Students who are diagnosed with ADHD can receive academic accommodations (e.g., private testing environments, extended test time, and alternative courses) that can significantly improve their grades (Harrison, Edwards, & Parker, 2007). Medications prescribed to treat ADHD are often viewed as tools to enhance studying and testing performance in the normal population, and are illicitly purchased and sold for these purposes (Barrett, Darredeau, Bordy, & Pihl, 2005). Stimulants used for the treatment of ADHD can be used as relatively inexpensive alternatives to cocaine or methamphetamine to get high (Swanson & Volkow, 2003).

Unfortunately, for an individual seeking to obtain an ADHD medication prescription for these inappropriate uses, it is quite easy to feign ADHD symptoms during a typical ADHD assessment. The symptoms of ADHD are widely publicized in the popular press and on the Internet and, consequently, are quite well known by the general public (Conti, 2004). Tan, Slick, Strauss, and Hultsch (2002) reported that malingerers spend a significant amount of time reading relevant literature concerning the diagnosis they are attempting to feign, scanning the Internet for information regarding that illness, and even getting ideas about the disorder from television shows. As a result, it certainly appears that individuals seeking to feign ADHD can readily educate themselves to the point that they can appear to meet ADHD diagnostic criteria when completing behavior rating scales and answering clinical interview questions, and can exhibit ADHD behaviors (e.g., restlessness) during the relatively short interval during which they are observed during an assessment.

It is well understood that the ADHD behavior rating scales are highly vulnerable to non-credible symptom reporting. Quinn (2003) found that the ADHD Behavior Checklist-Current scale (Murphy & Barkley, 1996) did not discriminate between college students asked to simulate ADHD and those diagnosed with ADHD. In fact, the mean scores for the diagnosed ADHD group did not fall in the clinical range according to DSM criteria, while the mean scores of those instructed to simulate ADHD did.

Harrison et al. (2007) compared the manner in which 72 patients previously diagnosed with ADHD and 70 university undergraduate students recruited for the study completed the Conners Adult ADHD Rating Scale (Conners, Erhardt, & Sparrow, 1999). The 70 students were divided into two groups: a group instructed to fake ADHD symptoms and given some general information regarding the DSM-IV criteria for ADHD, and a group instructed to simply respond honestly. Harrison et al. found that the faking group actually had higher scores on the CAARS

DSM-IV inattentive symptoms, hyperactive/impulsive symptoms, and ADHD symptoms total scales as well as the ADHD index than the patients previously diagnosed with ADHD.

In a slightly different research design, Fisher and Watkins (2008) asked 175 college students without a history of ADHD to complete the Barkley ADHD Behavior Checklist-Current Scale ($n = 88$) and the College ADHD Response Evaluation ($n = 87$) (Glutting, Monagan, Adams, & Sheslow, 2002) scales respectively. Participants were instructed to complete the scales in such a manner as to obtain a diagnosis of ADHD. They were then presented with the DSM-IV ADHD criteria and given 5 minutes to study them. After being warned that a professional might detect faking by looking for over-endorsement of symptoms, they then completed the behavior rating scales. Ultimately, 93% were able to successfully fake ADHD on the CARE and 77% successfully faked ADHD on the ADHD Behavior Checklist.

Despite the now well-recognized fact that it is easy to simulate the disorder on ADHD symptom report measures, the most widely used adult ADHD behavior rating scales fail to include any scales or other indicators to identify symptom exaggeration. The Barkley Adult ADHD Self-Report Forms for both current symptoms and childhood symptoms (Barkley & Murphy, 2006) and the Brown Attention-Deficit Disorder Scale (Brown, 1996) have no scales to identify any form of invalid reporting on these instruments. The Conners (CAARS) behavior rating scale does contain such a scale. However, it is an inconsistency scale designed only to detect contradictory endorsement of essentially the same ADHD symptoms. A high score on this scale only raises the possibility that the patient has completed the CAARS in an inaccurate manner. This could be due to failure to pay close attention to the content of the CAARS because of legitimate confusion and attention deficits, to impaired reading or verbal comprehension, or to lack of motivation. However, the CAARS inconsistency scale certainly does not directly address the issue of symptom exaggeration. The only adult ADHD self-report behavior rating scale that includes an index to detect symptom exaggeration is the Clinical Assessment of Attention Deficit-Adult (CAT-A) Infrequency Scale (Bracken & Boatwright, 2005). Unfortunately, the effectiveness of this scale in detecting symptom exaggeration is largely unknown, for there are no published research studies addressing this issue beyond the few studies conducted by the developers of the scale.

Very little research has addressed the effectiveness of embedded and free-standing SVT measures in identifying suspect effort in cognitive test performance in individuals referred for assessment of ADHD. The research on the effectiveness of measures designed to identify invalid completion of self-report ADHD behavior rating scales is also limited. In addition, there are very few data regarding the base-rate of symptom exaggeration and suspect effort in self-referred adults other than college students presenting for assessment of ADHD.

Only two major studies have been conducted examining the ability of free-standing and embedded SVT measures to detect suspect effort in actual patients undergoing ADHD evaluations. Sullivan, May, and Galbally (2007) studied the usefulness of the Word Memory Test (WMT) (Green, 2003) in identifying symptom exaggeration in 66 college students presenting for evaluation of ADHD, LD, or both ADHD and LD concurrently. They found that 47.6% (i.e., 10 of 21) of the students who were presenting for the assessment of only ADHD scored below the recommended cutoff scores on one or more of the WMT effort measures. In contrast, the rates of failure on the WMT were 15.4% (2 of 13) and 9.4% (3 of 32) for students presenting for assessment of only LD and concurrent LD and ADD respectively. The authors did not report the sensitivity and specificity of the WMT in identifying suspect effort in ADHD assessment. The small size of the group presenting for ADHD assessment only ($n = 21$), the fact that all participants were college students, and the limited number of symptom validity measures (i.e., only the WMT) are significant limitations of this study.

Suhr, Hammers, Dobbins-Buckland, Zimak, and Hughes (2008) attempted to replicate and extend the findings from the Sullivan et al. study (2007) by examining the effectiveness of not only the WMT but five additional SVT measures in identifying suspect effort in ADHD evaluations. These were: Exaggeration Index of for an expanded Auditory Verbal Learning Test (EIAVLTX; Barrash, Suhr, & Manzel, 2004), WAIS-III digit span less than 5 (Iverson & Franzen, 1994), WAIS-III Working Memory Index less than 70 (Etherton, Bianchini, Ciota, Heinly, & Greve, 2006), AVLT recognition less than 10 (Boone, Lu, & Wen, 2005), and WAIS-III Vocabulary-Digit span score of 2 or greater (Greve, Bianchini, Mathias, Houston, & Crouch, 2003). The study sample was drawn from individuals who had undergone neuropsychological assessment at a university psychology clinic. The 26 patients—31% of the initial study participants—who failed any one of the four subtests of the WMT were assigned to the non-credible effort group. This was the only criterion by which patients were assigned to the non-credible effort group. In addition, 15 patients were diagnosed with ADHD by multiple criteria and 24 patients were deemed to have significant psychological symptoms but were not diagnosed with ADHD. Suhr et al. (2008) found that, while possessing very high specificity, four SVT measures had very poor sensitivity in identifying non-credible effort. The sensitivity of the EIAVLTX was 8%, of Digit Span 19%, of the Working Memory Index 4%, and of the AVLT recognition measure 12%. The Vocabulary-Digit Span cutoff had a stronger sensitivity of 31% but an unacceptable specificity of 85%.

Only two studies have examined the ability of response bias scales or scores on clinical questionnaires alone to detect invalid completion of self-report symptom measures in actual patients undergoing ADHD evaluations. In the study described immediately above, Suhr et al. (2008) examined the ability of the CAARS inconsistency score to discriminate between the non-credible group, the ADHD group, and the psychological problem group. Scores above the clinical cutoff on the inconsistency scale identified only 3 of the 26 patients (i.e., 12%) in the non-credible group. No scores on any CAARS subscales discriminated between the non-credible group and the ADHD group.

Sullivan et al. (2007) examined the utility of the three negative response bias scales of the Personality Assessment Inventory (Morey, 1991) in detecting faking of ADD symptoms. Research has shown that the negative impression, malingering, and Rogers discriminate function scales of the PAI are effective in identifying simulators instructed to feign psychopathology (e.g., major depression) in college populations (Morey, 1996). As the authors had hypothesized, these three scales were not useful as very few of the individuals in the non-credible group had significantly elevated scores on these PAI response bias scales.

In conclusion, it is highly likely that at least some individuals undergoing adult ADHD evaluations will make a suspect effort in cognitive testing given the rate of malingering identified in other clinical populations and the substantial incentives for seeking a diagnosis of ADHD. It is also clear that it is easy for individuals to educate themselves so that they can present with symptoms consistent with ADHD in completing behavior rating scales and in answering questions in a clinical interview. Therefore there is a major need to identify measures that are effective in identifying suspect effort in testing and in completing behavior rating scales in this population. Unfortunately, the research to date addressing this need is small and has significant limitations.

The two studies (Suhr et al., 2008; Sullivan et al., 2007) both classified individuals as making a suspect effort solely on the basis of their failing one SVT measure (the WMT) and studied only college students. The Sullivan et al. study examined a very small group presenting for ADHD ($n = 21$). The Suhr et al. study examined the effectiveness of four SVT measures with demonstrated sensitivity in identifying suspect effort in other clinical populations. However, these four SVT measures are embedded in tests of memory, verbal attention capacity, working memory, and vocabulary. Their very poor sensitivity in identifying suspect effort in ADHD evaluations may be the result of their not appearing to be tests of sustained attention. A malingerer logically could assume they should feign deficits on tests of sustained attention; this is the hallmark cognitive deficit of this disorder. Finally, the limited research indicates that the CAARS inconsistency score and PAI negative response bias scales are all ineffective in detecting invalid completion of self-report symptom measures in ADHD evaluations.

To address some of the limitations of this prior research, the battery of neuropsychological tests employed in the current study includes six different embedded and free-standing symptom validity measures designed to detect suspect effort (i.e., exaggeration or feigning of cognitive deficits). No studies have examined the effectiveness of these widely used SVT measures in identifying suspect effort during cognitive testing in ADHD evaluations. This study is also the first to evaluate the effectiveness of several TOVA variables—including the new embedded symptom validity measures—as well as Conners CPT-II measures in identifying suspect effort in a non-credible effort group of patients referred for assessment of adult ADHD. Among the multiple ADHD behavior rating scales employed in the current study, we included an embedded validity scale (the CAT-A Infrequency Scale) consisting of behaviors that resemble ADHD symptoms but in fact are infrequently endorsed by patients with ADHD. This allowed the authors to identify invalid responding on behavior rating scales, thereby helping to ensure the accuracy and validity of the final diagnosis of ADHD. Finally, the study examined the behavior of real world patients referred for adult ADHD evaluations across a wide age range (not just college students).

These features of the present study will hopefully provide information useful in identifying symptom exaggeration among patients undergoing ADHD evaluations. The improved diagnostic accuracy of both genuine and feigned ADHD should also lead to the improved characterization of the cognitive deficits associated with ADHD in adults. Inclusion of these symptom validity measures will also allow us to better identify which of the SVTs employed are most effective in identifying suspect test taking effort in this population.

The primary purposes of this study were to (1) determine the relative efficacy of several free-standing and embedded symptom validity test measures in identifying suspect effort in cognitive testing during ADHD evaluations and (2) to determine the relative utility of two measures in identifying invalid completion of ADHD behavior rating scales. In addition, we sought to provide information on the sensitivity and specificity rates as well as positive predictive accuracy and negative predictive accuracy of a wide range of measures of potential utility in adult ADHD evaluations. Recommendations about the usefulness and validity of various cutoff scores on these symptom validity measures are provided.

Research and clinical experience suggests that malingerers often feign cognitive deficits by intentionally executing tasks at a very slowed speed. Patients malingering deficits on neuropsychological assessments conducted in the context of litigation and disability evaluations are often significantly slower in counting dots on the Dot Counting test (Boone, Lu, & Herzberg, 2002a), in circling the letter b on the b Test (Boone et al., 2002b), in recalling numbers on digit span forward (Babikan, Boone, Lu, & Arnold, 2006), in finger tapping (Arnold et al., 2005), and in completing the recognition trial of the Warrington Recognition Memory Test – Words (Kim et al., 2010). Furthermore, when asked to identify the strategies they employed to fake ADHD, 31% of the students instructed to feign ADHD in a simulation study said they did so by completing tasks slowly (Harrison et al., 2007). Therefore we examined the effectiveness of several tests with time to completion components in identifying suspect effort in our population of adults undergoing ADHD assessment.

In addition, this study sought to address a major research question that has not been studied sufficiently in research on ADHD to date. That is, whether and to what degree failures on specific SVTs are associated with poorer performance on tests in the same cognitive domain and unrelated cognitive domains in neuropsychological tests commonly employed in the assessment of ADHD. Failure on a SVT based in a recognition memory test (the Test of Memory Malingering, Tombaugh, 1996) has been found to predict lower scores on standardized intellectual and neuropsychological tests—not just tests of recognition memory or memory per se—in litigants with mild traumatic brain injuries (Constantinou, Bauer, Ashendorf, Fisher, & McCaffrey, 2005). Furthermore, Green, Rohling, Lees-Haley, and Allen (2001) contend that the performance on the Word Memory Test accounted for approximately 50% of the variance in a composite score derived from 43 different neuropsychological tests in a large sample of patients with mixed diagnoses but being evaluated in the context of compensation litigation.

Only Sullivan et al. (2007) have studied the relationship between failure on a SVT and performance on cognitive testing in ADHD. They found that performances on all three WMT effort test measures (i.e., immediate recognition, delayed recognition, and consistency scores) were significantly correlated with Full Scale IQ scores and two California Verbal Learning Test-2 (CVLT-II) measures (i.e., initial free recall and long delay free recall) for students referred for ADHD evaluation only. However, the WMT was the only SVT measure and the CVLT-II was the only cognitive measure in this small study of college students presenting for ADHD evaluation.

Finally, this study also addressed another important but very little studied issue. That is, whether and to what degree is symptom exaggeration on ADHD self-report behavior rating scales associated with (a) suspect effort on SVT measures related to cognitive testing and (b) poorer performance on standard cognitive tests often employed in the assessment of adult ADHD? Sullivan et al. (2007) found that performances on all the three WMT effort test measures were significantly correlated with the Conners Adult ADHD Rating Scale, Long Version (Conners et al., 1999) DSM-IV ADHD hyperactive/impulsive scale and the DSM-IV ADHD symptoms total scale measures for college students referred for ADHD evaluation only.

METHOD

Participants

This archival study consists of 326 patients. All study patients were referred for an ADHD assessment in a general neuropsychological practice in the Department of Psychiatry at Hennepin County Medical Center in Minneapolis, Minnesota. Virtually all of these patients were outpatients rather than inpatients. These assessments occurred between July 2005 and February 2010. Of these patients, 268 were all the consecutive referrals for assessment of ADHD who were not disqualified by exclusion criteria. Virtually all of these patients were also initially self-referred to other medical professionals and then re-referred to our neuropsychology clinic by these professionals. They were physically healthy and faced with potentially receiving a diagnosis of ADHD that they perceived as probably being beneficial to them in many ways whether they truly had ADHD or not. Consequently, there was little reason to suspect that they might make a less than an adequate effort in testing unless they were, consciously or otherwise, attempting to feign symptoms or exaggerate any real cognitive deficits. The majority of analyses in this study focused on this group of 268 patients referred for ADHD assessment.

A total of 58 patients who were referred and assessed for ADHD in this time period were excluded from this study. They were excluded because they had neurological conditions, head injuries, learning disabilities, mental retardation, a psychiatric disorder other than depression or anxiety, major substance abuse/dependence, or a physical illness at a level of severity that could have caused significant cognitive deficits, including those cognitive deficits associated with ADHD.

A significant percentage (55%) of the 268 patients reported experiencing a mood disorder: 24% were diagnosed with depression NOS, 12% were diagnosed with major depression, 9% with mood disorder NOS, and 10% with anxiety disorder NOS. In addition, 39% of the patients were taking an anti-depressant medication while 8% were taking an anxiolytic medication. Finally, 21%, 17%, and 12% of the patients reported using alcohol, marijuana, and cocaine excessively respectively. These percentages of individuals with these co-morbid psychiatric conditions are similar to those found in prior ADHD research (Barkley, Murphy, & Fischer, 2008; Wilens et al., 2009). All of these patients had not been excluded because their frequency of use, number of years of use, amount of the substance consumed per week, and recency of use of these substances were not considered to be capable of causing impaired cognitive test performances at the time the ADD assessments were conducted.

Patients with an estimated WAIS-III Full Scale IQ score of less than 70 were excluded from the study given the now well-understood fact that many SVTs have unacceptable specificity rates and are therefore not valid for use with individuals in the "mentally retarded" or extremely low range of functioning (Dean, Victor, Boone, & Arnold, 2008; Marshall & Happe, 2007). The demographics of the study patients are presented in Table 1. It should also be noted that 74% of the sample was aged 17–30. Of the 268 patients, 151 (56%) were attending college at the time they underwent ADHD evaluation. Only 11% of the patients had less than average range estimated Full Scale IQ scores; 4% were in the borderline range; 41% of the patients had high average or better estimated Full Scale IQ scores.

Procedures

The cognitive tests used in these neuropsychological assessments were administered according to standardized instructions by either a board-certified neuropsychologist or by trained clinical psychology doctoral students working under the supervision of a board-certified neuropsychologist. Patients who were prescribed medications for the treatment of ADHD had not taken these medications for a minimum of 24 or 48 hours before testing for short acting and long acting preparations respectively (Nigg, 2005).

Consistent with the recommendations of a multi-method approach to the assessment of adult ADHD (Frazier, Demaree, & Youngstrom, 2004; Hervey, Epstein, & Curry, 2004), all study patients ($n = 268$) completed thorough

Table 1 Demographic characteristics and mean estimated IQ scores

Demographics	*N*	% of sample
Ethnicity/Race		
Caucasian	218	81
African-American	24	9
Hispanic, Asian, Native American, or other race	26	10
Sex		
Male	163	61
Female	105	39
	M (*SD*)	Range (min–max)
Age (years)	27.8 (9.1)	17–55
WAIS-III estimated Full Scale IQ	112 (17.0)	71–149
Education	h.s. graduate + 1 year college	7th grade-graduate student

neuropsychological assessments. The assessments included numerous cognitive tests, an extensive clinical interview, behavioral observations, standardized self-report instruments (i.e., ADD behavior rating scales completed by both the patient and another individual who knew him well), and effort testing. All patients received at least the same core cognitive test battery.

The core test battery included the vocabulary, comprehension, similarities, block design, symbol search, digit symbol, matrix reasoning, and letter–number sequencing subtests of the Wechsler Adult Intelligence Scale-III (WAIS-III) (Wechsler, 1997). The core battery also included the Rey Complex Figure (Meyers & Meyers, 1995), the California Verbal Learning Test-II (CVLT-II) (Delis, Kramer, Kaplan, & Ober, 2000), the Verbal Fluency, Design Fluency, Color-Word Interference, and Tower Tests from the Delis-Kaplan Executive Function System (DKEFS) (Delis, Kaplan, & Kramer, 2001), the NAB Number and Letters Test (White & Stern, 2003), the Short Category Test (Wetzel & Boll, 1991), the Paced Auditory Serial Addition Test (PASAT)-100 (Diehr et al., 2003), the Salthouse Reading Span Task (Salthouse, 1994), the Dot Counting Test (Boone et al., 2002a), the Sentence Repetition Test (Strauss, Sherman, & Spreen, 2006), and the b-Test (Boone et al., 2002b). Of the study patients, 38% ($n = 101$), who were assessed prior to 2008 were also given the Conners Continuous Performance Test-II (CCPT-II) in addition to the standard battery noted above (Conners, 2000). The great majority of the remainder of the patients in this study ($n = 167$), who were assessed in 2008 through 2010 were given the Test of Variables of Attention (TOVA) (Greenberg, Kindschi, Duprey, & Corman, 1996) as well as the core test battery.

The specific embedded and free-standing SVT measure indices and their respective cutoff score for initially identifying suspect effort included in the test batteries administered to all 268 study patients were the following: a score of 6 or less on the Reliable Digit Span score (Babikian et al., 2006), 2 or more errors on the CVLT-II Forced Choice Recognition Test (Root, Robbins, Chang, & Van Gorp, 2006), a Sentence Repetition score of 10 or less (Schroeder & Marshall, 2010), a Dot Counting Test e score of 17 or greater (Boone et al., (2002a), and a b-Test e score of 120 or more, 5 or more commission errors, 2 or more d errors, and time to completion of 600 or more seconds (Boone et al., 2002b). The TOVA SVT cutoff scores used to initially identify suspect effort were the following recommended by the TOVA corporation: total response time variability > 180 milliseconds, and an omission error standard score < 45 (Hughes, personal communication). Finally, a subset of the sample (20%) of the study patients were given the Word Memory Test (WMT) (Green, 2003) and the authors used the standard cutoff scores ($<82.5\%$ correct) for immediate recognition, delay recognition, and recall consistency scores in identifying suspect effort. Patients were given only the immediate and delayed recognition trials of the WMT. Consequently, the additional profile analyses that have been helpful in identifying suspect effort in other patient populations could not be conducted.

These embedded and free-standing SVT measures were interspersed throughout the entire test battery, thereby enabling the continuous monitoring of the patient's effort over the course of a testing session. As Boone (2009) has noted, such continuous monitoring is important because an individual's effort can fluctuate significantly over the course of an assessment for several reasons. Using several

SVTs in a test battery also enables the clinician to evaluate the patient's effort on tests involving several different cognitive abilities, e.g., visual and verbal attention, working memory, memory, executive functions, and speed of cognitive processing. This advantage of using several SVTs is also important because some individuals exaggerate or feign deficits in only one or two specific cognitive abilities, typically the cognitive abilities they complain of having difficulty with (Boone, 2009).

A total of 232 patients (87% of the study sample) given the TOVA or the CCPT as well as the core cognitive test battery also completed the Barkley Adult ADHD Self-Report Forms (BAASRF) for both current symptoms and childhood symptoms (Barkley & Murphy, 2006). Both the current (adult) and childhood symptom forms include 18 behavioral symptoms. Nine of these behavioral symptoms that refer to inattention are alternated with nine behavioral symptoms that refer to hyperactive-impulsive behavior. For each behavioral symptom, the patient indicates whether they exhibit that behavior "rarely or never" (scored 0), "sometimes" (scored 1), "often" (scored 2), or "very often" (scored 3).

Whether the patient is endorsing experiencing symptoms consistent with ADHD is determined in two ways. First, the sum of the scores for all nine inattention symptoms and for all nine hyperactivity/impulsivity symptoms are compared to age- and sex-appropriate cutoff scores. For example, if a 20-year-old patient has a total inattention symptom score of 14, he/she would be considered to meet criteria for having inattention problems consistent with the diagnosis of ADHD. Second, each item that is reported as being exhibited "often" or "very often" is considered to be an endorsement of that symptom, and counts as one symptom toward meeting the six out of nine symptoms necessary on either the inattention or hyperactivity lists to meet DSM-IV criteria for the diagnosis of ADHD. Thus both the current (adult) and childhood scales produce four scores: (a) a sum of the item scores for inattention, (b) a sum of the item scores for hyperactivity-impulsivity, (c) a total inattention symptom count, and (d) a total hyperactivity-impulsivity symptom count. We also derived a single score for each of these four scores indicating whether or not the patient met the sum of the inattention item scores, sum of the hyperactivity-impulsivity item score, total inattention symptom count, and total hyperactivity-impulsivity symptom cutoff scores for diagnosis of ADHD.

Of these 232 patients, 167 (62% of the study sample) were also given two supplemental behavior rating scales for their current (i.e., adult) symptoms. The first was the nine behavioral symptoms that Barkley has identified as being most effective in distinguishing patients with ADHD from community control and clinical control participants in his research (Barkley et al., 2008, p. 437). Barkley has concluded that an individual who endorses experiencing six or more of these nine symptoms should be diagnosed with ADHD. The second supplemental scale was the 10 behavioral items composing the Clinical Assessment of Attention Deficit-Adult (CAT-A) Infrequency Scale (Bracken & Boatwright, 2005, p. 29). A typical neuropsychological assessment behavior observation sheet consisting of visual analog scales of 39 behavioral dimensions was completed by psychometrists conducting testing on all 268 patients in the study. Of the 39 behavioral dimensions, psychometrists were asked to rate the degree to which the patient exhibited behavior on visual analog scales concerning three behaviors of particular relevance for the diagnosis of ADD: a motor activity dimension ranging from the anchors "calm" to "restless," a spontaneous conversation dimension ranging from "appropriate" to "excessive," and an attention dimension ranging from "attentive" to "distractible."

The individual's ratings on these three visual analog scales were compared to the frequency with which 232 study patients endorsed experiencing three behaviors on the Barkley Adult ADHD Self-Report Forms (BAASRF) for current symptoms: "feel restless", "talk excessively," and "am easily distracted." There had to be the most extreme discrepancy possible for at least two of the three behavioral dimensions for the overall discrepancy between the patient's ratings of behaviors on the BAASRF and the psychometrist's behavior observation ratings to be suggestive of invalid completion of the self-report behavior rating scales. For example, a patient's response on the behavior rating scale regarding the attention dimension was considered an extreme discrepancy only if they reported being "easily distracted" "very often" on the BAASRF, while the psychometrist also rated them as exhibiting no distractibility whatsoever during testing on the behavior observation sheet visual analog scale for the "attentive–distractible" attention dimension. Completion of the 10 behavioral items comprising the CAT-A Infrequency scale reveals the degree to which the patient "strongly agrees" that they experience certain behaviors that were endorsed in such a manner by only a few (usually less than 2%) of the participants in the CAT standardization sample (Bracken & Boatwright, 2005, p. 29). Therefore, endorsement of these behaviors raises the possibility that the patient is exaggerating ADHD symptomatology beyond the norm even for those individuals with the disorder.

For example, very few participants checked off "strongly agree" with respect to their exhibiting the behavior "I work more energetically than most people". Only 1% of the entire CAT-A standardization sample and 6% of the combined clinical sample indicated that they "strongly agree" that they experienced 3 of the 10 behaviors comprising the Infrequency scale (Bracken & Boatwright, 2005, p. 35). Therefore, for the purposes of the current study, the CAT-A Infrequency scale responses were thought to suggest invalid completion of the self-report behavior rating scales only if the patient endorsed experiencing "very often" on at least 3 of the 10 items on the Infrequency scale.

Patients were diagnosed as having made a suspect effort in cognitive testing and/or invalid completion of the self-report ADD behavior rating scales and were placed in the non-credible group if they met the set of criteria stipulated in either condition I or II (see Table 2). Note that criteria I and II both meet the Slick, Sherman, and Iverson (1999) criteria for probable malingered neurocognitive dysfunction, i.e., external incentive plus two or more effort test failures (excluding below-chance performance on forced choice tests) or one effort test failure and one instance of non-credible self-reported symptoms or test data discrepancy.

Failing two separate SVT measures or failing one SVT measure as well as rendering an unusually impaired performance on one cognitive test are both very low-probability events that are highly likely to be indicative of suspect effort. In their study of the effectiveness of four embedded SVTs, Victor, Boone, Serpa, Buehler, and Ziegler (2009) found that failure of any two embedded SVTs had a

Table 2 Criteria for having made suspect effort

CONDITION I
There was failure on two separate SVT measures (e.g., the b Test and the Dot Counting Test)
OR
There was failure on one SVT measure (e.g., b Test) *and* an unusually impaired performance (two standard deviations or more below the average of ADD patients) on one cognitive test (e.g., the CCPT omission errors or the learning trials of the CVLT-II test).
CONDITION II
There was failure on one SVT measure *or* an unusually impaired performance on one cognitive test
AND
There was an invalid completion of behavior rating scales indicated by a CAT-A Infrequency scale score of 3 or higher *or* a significant discrepancy between the patient's ratings of their distractibility, restlessness, and excessive spontaneous conversation on the Barkley Adult ADHD Self-Report form and the psychometrist's ratings concerning those behaviors in their general behavioral observations.

sensitivity of 83.8% and specificity of 93.9% in identifying suspect effort in testing. Similarly, Larrabee (2003) found that failure on any two of five embedded SVTs had a sensitivity of 87.5% and specificity of 94.4% in identifying suspect effort in his combined derivation and cross-validation samples.

Patients were diagnosed as having adult attention deficit disorder if they met all four of the following criteria: (a) they had completed behavior rating scales in a valid manner and scale scores were indicative of adult ADHD, (b) they had reported significant impairment in at least one major setting (educational, occupational, home life/family function) in an extensive clinical interview, (c) they had reported the presence of some ADHD related symptoms impairing performance before age 16 that were not plausibly accounted for by a comorbid disorder (depression, anxiety) in the clinical interview, and (d) they had at least some valid cognitive testing results that were indicative of ADHD, at minimum impaired sustained attention on at least one test. There were some patients who clearly met the adult ADHD diagnostic criteria but also gave suspect efforts on SVT measures and invalid completion of ADD behavior ratings scales. These were individuals in whom the weight of evidence strongly suggested that they had ADHD, but apparently they felt they needed to insure they would receive the diagnosis by exaggerating some cognitive deficits and behavioral problems.

Data analyses

The fact that multiple failures of free-standing or embedded symptom validity tests have such good sensitivity and specificity in identifying suspect effort was the primary basis for classifying patients in this study as giving a suspect effort or not. However, as Nelson et al. (2003) have pointed out, it is critical to examine the inter-test correlations between measures of suspect effort to insure that they are not highly correlated with one another and, thereby, represent relatively independent sources of information with respect to test-taking effort. Using highly correlated SVT measures diminishes the likelihood of detecting suspect effort (Rosenfeld, Sands, & Van Gorp, 2000).

To address this concern the authors first identified a single score that best represented each individual SVT measure. We assumed that scores drawn from the same test would be highly correlated. Consequently, only correlations between scores across tests were examined. The scores chosen were as follows: CVLT forced choice recognition errors, b Test e score, dot counting test e score, sentence repetition score, reliable digit span score, word memory test consistency score, the CAT-A Infrequency scale score, TOVA omission errors score, and CCPT omission errors. Only the sentence repetition score and reliable digit span scores had normal distributions, and a Pearson zero-order correlation was run to examine their relationship. Spearman's rho correlations were run to examine the relationships between all other pairs of SVT measures when at least one of the SVT measures was not normally distributed. An adjusted alpha of $p < .01$ was used for all significance testing because of the large number of correlations run.

Frequency analyses were conducted once the 268 participants were divided into those who were and were not in the non-credible group. These analyses were conducted to determine the rates at which patients in these two groups passed or failed the initial cutoff scores for suspect effort for the Reliable Digit Span, the CVLT-II Forced Choice Recognition Test, the Sentence Repetition Test, the Dot Counting Test, b Test, TOVA test, and Word Memory Test indices. Additional frequency analyses were conducted to identify the rates at which the participants in both the credible and non-credible effort groups passed or failed the following test measures: two new TOVA effort indices (commission error reaction time > overall reaction time and post commission error speed < overall reaction time speed), the two CCPT effort measures, the CAT-A Infrequency Scale, and the discrepancy between the patient ratings of behaviors on the Barkley Adult ADHD Self-Report form and the psychometrist's behavior observation ratings.

It is widely accepted that a symptom validity test measure must have a cutoff score with a specificity rate above 90% for it to be considered a valid clinical measure of suspect effort (Boone, 2007). Therefore, if the initial cutoff scores did not have a specificity rate exceeding 90%, additional analyses were conducted to identify cutoff scores for the suspect effort indices that resulted in the optimal combination of sensitivity and specificity rates. Finally, once sensitivity and specificity rates were determined for each suspect effort measure in the ADHD neuropsychological assessment, these figures were used to calculate the positive predictive accuracy and negative predictive accuracy rates for identifying suspect effort in each group of patients in this study.

Following the Grimes and Schulz (2005) method as applied by Larrabee (2008), likelihood ratios (LRs) were used to examine the relationship of multiple SVT failure to the probability of malingering. LRs are defined as the ratio of sensitivity to 1 minus sensitivity (i.e., sensitivity/1 – sensitivity) and, in this context, represent the likelihood of a positive test result in someone who is malingering compared to the likelihood of a positive test result in someone who is not malingering. Post-test odds of malingering can be computed by multiplying the LR by pre-test odds of malingering (base rate/1 – base rate). The post-odds can then be converted to a post-test probability of malingering using the formula (odds/odds + 1). Importantly, if the diagnostic indicators are independent—which the SVT measures are in this study—the post-test odds for one SVT measure can be used as the pre-test odds for a second SVT measure, thereby "chaining" the likelihood ratios. The six most sensitive SVT measures were included in determining the post-test probability of failing any combination of two SVT measures and of failing any combination of three SVT measures. The six most sensitive SVT measures were the following: WMT immediate recall score, b Test e score, TOVA reaction time variability, TOVA omission errors, CAT-A self-report Infrequency scale score, and the discrepancy between the patient's ratings of behaviors on the Barkley Adult ADHD Self-Report Form and the psychometrist's behavior ratings.

Previous research has suggested that individuals feigning ADHD can successfully fake ADHD symptoms on behavior rating scales. To address this issue once again, one-way ANOVAs were run to determine if there were significant differences between the ADHD credible, non-ADHD credible, ADHD suspect, and the non-ADHD suspect groups on eight Barkley Adult ADHD Self-report Form variable scores as previously discussed. The authors also compared the four groups with respect to their endorsement of the nine symptoms that Barkley has recently identified as being most effective in distinguishing patients with ADHD from community control and clinical control participants. Post-hoc *t*-tests with Bonferroni corrections were run to determine if there were significant differences on each of these nine ADHD behavior rating scale variables between each of the four patient groups. Chi-square tests were also run to determine if there were significant differences with respect to the percentage of patients within each group whose scores on the same eight Barkley Adult ADHD Self-report Form variable scores and the new Barkley nine symptom scale met the cutoff score to be considered indicative of the individual having ADHD.

To investigate whether those feigning ADHD feign cognitive tests by executing tasks in a slowed manner, *t*-tests were run comparing the performance of patients identified as making a non-credible effort versus those making a credible effort on eight tests that measure the cognitive processing speed with which a task is executed. These tests were the WAIS-III Symbol Search test, the WAIS-III Digit Symbol test, the DKEFS Color Word Interference Test (DKEFS-CWIT) combined naming and reading score, the Conners CPT (CCPT) reaction time score, and the NAB Number and Letter Test part A (NAB-NLA) speed score. Next, frequency analyses identified the rates at which the credible and non-credible effort patients passed or failed these test measures. Further analyses identified cutoff scores for these test measures that resulted in the optimal combination of sensitivity and specificity rates.

Finally, statistical analyses were conducted to address the issue of whether and to what degree failures on specific SVT measures are associated with poorer performance on tests in the same and other cognitive domains. T-tests were run comparing the performance of the study patients passing and failing each of nine different SVT measures on 44 different cognitive test measures. The initial *t*-tests indicated that those failing the nine SVT measures had statistically significant poorer performances on the vast majority of these 44 measures, including estimated WAIS-III Full Scale IQ score. When ANCOVAs were run with FSIQ as the covariate, those failing the SVT measures had statistically significant poorer performances on a much smaller number of cognitive test measures.

RESULTS

The correlational relationships between the SVT measures are summarized in Table 3. A total of 20 correlations were significant at the $p < .01$ level. However, none of these pairs of SVT measure variables had a shared variance of 50%. In fact, the highest correlation was between the reliable digit span score and the sentence repetition score ($r = .539$) with a shared variance of 29%. Only two other pairs of measures had a shared variance as high as 16%: the dot counting e score and b Test e score (rho = .411) and the dot counting e score and the reliable digit span score (rho = .426). As Nelson et al. (2003) have argued, their low shared variance suggests these SVT measures represent relatively independent measures of effort and, therefore, the concurrent use of these SVT measures in diagnosing suspect effort is appropriate.

A total of 59 patients (22% of the 268 study participants) were identified as having made a suspect effort in cognitive testing and/or in completion of the self-report ADHD behavior rating scales and were placed in the non-credible groups based on the diagnostic criteria previously discussed. A total of 80 patients (30% of the 268 study participants) were diagnosed with ADHD based on the diagnostic criteria noted above. Of these, 29 were diagnosed with ADHD combined type, 36 were diagnosed with ADHD inattentive type, and 15 were not diagnosed by subtype because of insufficient information. The diagnosis of ADHD combined type versus inattentive type was based on their completion of the Barkley Adult ADHD Self-report form current and childhood inattention and hyperactive/impulsive items.

Of these 80 individuals diagnosed with ADHD, 66 (83%) gave a credible effort in cognitive testing and in completing behavior rating scales and are considered the ADHD credible group. Of these 80 patients, 14 (17%) did not give a credible effort and are considered the ADHD suspect group. Of the 188 patients not diagnosed with ADHD, 143 (76%) gave a credible effort in cognitive testing and in completing behavior ratings scales and are considered the non-ADHD credible group. Of these 188 patients, 45 (24%) did not give a credible effort and are considered the non-ADHD suspect group. The means and standard deviations of the four diagnostic groups for the embedded and free-standing SVT measures are summarized in Table 4. Not surprisingly, on the vast majority of SVT measures, the non-ADHD suspect group performed more poorly than the ADHD credible and non-ADHD credible groups. The fact that the ADHD suspect group often did not have a statistically significant poorer performance on these SVT measures than the ADHD credible and non-ADHD credible groups is due to the small number (i.e., 14) patients in the ADHD suspect group.

As can be seen in Table 5, several SVT measures have good sensitivity and specificity in detecting suspect effort in cognitive testing. It is important to note that the WMT was introduced into the assessment battery late in this study and given to only 20% of the study participants. Measures to detect suspect completion of ADHD behavior rating scales also appear to be effective, although their specificities are slightly below standard.

Positive predictive accuracy and negative predictive accuracy rates are also reported in Table 5 for the two base rates that are most likely to be relevant to

Table 3 Correlations between symptom validity test measures

Measure	CCPT Omission	Dot Counting E- Score	Sentence Repetition	CVLT Forced Choice	TOVA Omission Errors	b Test E-score	Infrequency SS	Reliable Digit Span	Word Memory Test
CCPT Omission	**1.00**								
N	102								
Dot Counting E-score	.138	**1.00**							
N	100	298							
Sentence Repetition	−.105	−.362*	**1.00**						
N	102	298	302						
CVLT Forced Choice	−.217	−.134	.154*	**1.00**					
N	102	297	301	302					
TOVA Omission Errors	NA	.305*	−.292*	−.042	**1.00**				
N	NA	196	198	199	199				
b Test E-score	.181	.411*	−.320*	−.124	.373*	**1.00**			
N	94	283	287	288	193	288			
Infrequency SS	.866	.207*	−.235*	−.059	.235*	.317*	**1.00**		
N	3	195	197	198	196	194	198		
Reliable Digit Span	−.112	−.426*	.539*	.183*	−.325*	−.340*	−.224*	**1.00**	
N	102	297	301	302	199	288	198	302	
Word Memory Test Inconsistency Score	NA	−.338*	.357*	.030	−.241	−.329	−.108	.362*	**1.00**
N	NA	76	76	76	76	76	76	76	76

*$p < .01$.

Table 4 SVT measures means and standard deviations for each diagnostic group

Group category	ADD credible	ADD suspect	Non-ADD credible	Non-ADD suspect	Significance
Group number	1	2	3	4	–
Number of patients	66	14	143	45	–
Age	26.19 (8.43)	26.14 (7.61)	27.87 (8.81)	30.29 (10.42)	No Significance
FSIQ	117.51 (12.62)	108.57 (20.66)	115.15 (16.79)	98.09 (15.03)	1,3 > 4
CVLT-II Forced Choice	15.89 (0.31)	15.57 (0.76)	15.90 (0.32)	15.00 (2.43)	1,3 > 4
Sentence Repetition	14.78 (2.28)	14.57 (2.41)	15.28 (2.29)	13.11 (2.85)	1,3 > 4
Reliable Digit Span	10.07 (2.11)	8.86 (2.25)	10.42 (2.32)	8.76 (2.08)	1,3 > 4
Dot Counting E-score	9.58 (2.29)	11.21 (4.37)	8.75 (2.51)	12.93 (6.88)	1,3 < 4
b Test E-score	44.82 (14.68)	61.00 (20.69)	46.26 (48.02)	123.55 (117.18)	1,2,3 < 4
b Test d Error	0.17 (0.48)	0.93 (1.27)	0.11 (0.32)	1.64 (2.26)	1 < 4; & 3 < 2,4
b Test Commission Errors	0.48 (0.80)	1.00 (1.36)	0.95 (4.42)	5.84 (11.33)	1,2,3 < 4
b Test Omission Errors	11.61 (9.68)	13.86 (8.73)	8.23 (9.61)	21.32 (23.98)	1,3 < 4
b Test Total Time	406.23 (77.31)	468.07 (164.19)	412.13 (99.88)	479.25 (153.22)	1,3 < 4
WMT Immediate Recall	36.35 (3.15)	33.60 (3.29)	37.36 (3.26)	30.50 (3.99)	1,3 > 4
WMT Delayed Recall	38.20 (2.95)	37.20 (2.77)	38.05 (2.95)	32.50 (4.14)	1,3 > 4
WMT Consistency	36.85 (3.86)	34.00 (2.92)	36.91 (3.18)	29.83 (3.97)	1,3 > 4
TOVA RT Variability	138.82 (39.34)	172.00 (71.74)	102.09 (50.86)	186.63 (57.24)	3 < 1,2,4; & 1 < 4
TOVA Omission Errors	15.33 (30.17)	38.13 (43.67)	9.56 (24.86)	45.81 (50.89)	1,3 < 4
TOVA Commission Errors	15.49 (11.23)	22.38 (12.88)	10.49 (10.56)	28.41 (41.14)	1,3 < 4
Conner's Omission T Score	33.68 (17.40)	18.45 (21.35)	45.29 (15.57)	26.43 (23.75)	2,4 < 3
Conner's Commission T Score	38.33 (10.09)	32.17 (13.06)	45.98 (9.97)	43.74 (10.59)	1,2 < 3
CAT-A Self	1.18 (1.29)	3.33 (2.65)	0.93 (1.17)	2.97 (2.12)	1,3 < 2,4
CAT-A Other	1.15 (1.27)	1.33 (1.03)	1.13 (1.30)	1.62 (1.56)	No Significance

Table 5 Sensitivity, specificity, positive and negative predictive accuracy rates for SVT measures

Group	N	Cutoff	Sens. (%)	Spec. (%)	15% base rate PPA (%)	15% base rate NPA (%)	30% base rate PPA (%)	30% base rate NPA (%)
WMT Immediate Recall	53	≤82.5%	63.64	92.86	61.13	93.54	79.25	85.63
WMT Consistency	53	≤82.5%	63.64	95.24	70.23	93.69	85.14	85.94
WMT Pass/Fail	53	≤82.5%	63.64	90.48	54.12	93.38	74.13	85.31
TOVA Omission Errors	165	>25	62.86	91.54	56.73	93.32	76.10	85.19
CAT-A Self	167	3+	58.33	89.39	49.24	92.40	70.20	83.35
Conner's Omission T Score	101	< 20	56.52	87.00	43.41	91.90	65.08	82.36
TOVA RT Variability	165	>180	54.29	91.54	53.11	91.90	73.34	82.37
b Test E-score	257	70+	46.55	93.43	55.79	90.84	75.40	80.32
b Test E-score	257	90+	36.21	95.96	61.27	89.50	79.34	77.83
CAT-A Self	167	4+	36.11	96.97	67.77	89.58	83.63	77.98
BAASFR vs. Psychometrist's Ratings	249	2/3	34.62	86.29	30.83	88.21	51.97	75.49
b Test Commission Errors	257	2+	34.48	90.41	38.82	88.66	60.64	76.30
b Test Total Time	257	550+	34.48	92.93	46.25	88.93	67.64	76.80
Dot Counting E-score	263	14+	33.89	95.10	54.97	89.07	74.77	77.05
b Test d Error	257	2+	32.72	90.41	37.58	88.39	59.39	75.82
TOVA Commission Errors	165	>30	25.71	92.31	37.11	87.56	58.90	74.35
b Test Omission Errors	257	25+	24.14	91.92	34.52	87.29	56.15	73.87
b Test E-score	257	120+	22.41	97.47	60.99	87.68	79.15	74.56
b Test Total Time	257	600+	22.41	96.97	56.62	87.63	76.02	74.46
CAT-A Other	167	3+	22.22	86.41	22.39	86.29	41.20	72.16
Reliable Digit Span	267	≤7	22.03	92.31	33.58	87.03	55.11	73.42
CVLT-II Forced Choice Errors	267	2+	20.34	99.52	88.20	87.62	94.78	74.46
Dot Counting E-score	263	17+	20.34	99.51	87.99	87.76	94.68	74.46
WMT Delayed Recall	53	≤82.5%	18.18	90.48	25.21	86.24	45.01	72.07
b Test Commission Errors	257	5+	17.24	97.47	54.60	86.97	74.49	73.32
b Test Omission Errors	257	30+	17.24	94.44	35.37	86.61	57.06	72.70
Sentence Repetition	266	≤10	13.56	99.04	71.37	86.65	85.82	72.78
Reliable Digit Span	267	≤6	13.56	99.05	71.37	86.65	85.82	72.78
CAT-A Other	167	4+	7.41	94.17	18.32	85.21	35.26	70.35
Conner's Commission T Score	101	< 20	4.35	100.00	100.00	85.56	100.00	70.93

outpatient adult ADHD assessment settings: 15% and 30%. Consequently, the results of this study strongly suggest that, with appropriate cutoff scores, several SVT measures have impressive sensitivity for detecting suspect effort in individuals undergoing ADHD evaluations while also exhibiting appropriate specificity.

Table 6 depicts the post-test probabilities derived from the chaining of likelihood ratios for a patient making a suspect effort when he/she fails two and three SVT measures for each possible pair of the six most sensitive SVT measures and each possible triplet of these same SVT measures. Clearly there is an increase in the probability of an individual making a suspect effort as a function of the number of SVT measures failed. Failing three SVT measures generally increases the probability of suspect effort in comparison to failing two SVT measures. However,

Table 6 Probabilities of suspect effort derived from chaining of likelihood ratios for failures on two and three SVT measures at 15% and 30% malingering base rates

	Post-test probability 15%	Post-test probability 30%
TWO SVT FAILURES		
WMT/b Test E-score	0.949	0.979
WMT/TOVA Reaction Time Variability	0.944	0.976
WMT/TOVA Omission Errors	0.952	0.979
WMT/CAT-A Infrequency Scale – Self	0.936	0.972
WMT/BAASFR vs. Psychometrist's Ratings	0.870	0.942
b Test E-score/TOVA Reaction Time Variability	0.889	0.951
b Test E-score/TOVA Omission Errors	0.903	0.958
b Test E-score/CAT-A Infrequency Scale – Self	0.873	0.943
b Test E-score/BAASFR vs. Psychometrist's Ratings	0.759	0.885
TOVA Reaction Time Variability/CAT-A Infrequency Scale - Self	0.862	0.938
TOVA Reaction Time Variability/BAASFR vs. Psychometrist's Ratings	0.741	0.874
TOVA Omission Errors/CAT-A Infrequency Scale – Self	0.878	0.946
TOVA Omission Errors/BAASFR vs. Psychometrist's Ratings	0.768	0.889
CAT-A Infrequency Scale – Self/BAASFR vs. Psychometrist's Ratings	0.710	0.856

THREE SVT FAILURES		
WMT/b Test E-score/TOVA Reaction Time Variability	0.992	0.997
WMT/b Test E-score/TOVA Omission Errors	0.993	0.997
WMT/b Test E-score/CAT-A Infrequency Scale – Self	0.990	0.996
WMT/b Test E-score/BAASFR vs. Psychometrist's Ratings	0.979	0.996
WMT/TOVA Reaction Time Variability/CAT-A Infrequency Scale - Self	0.989	0.996
WMT/TOVA Reaction Time Variability/BAASFR vs. Psychometrist's Ratings	0.977	0.990
WMT/TOVA Omission Errors/CAT-A Infrequency Scale – Self	0.991	0.996
WMT/TOVA Omission Errors/BAASFR vs. Psychometrist's Ratings	0.980	0.992
WMT/CAT-A Infrequency Scale – Self/BAASFR vs. Psychometrist's Ratings	0.974	0.989
b Test E-score/TOVA Reaction Time Variability/CAT-A Infrequency Scale – Self	0.978	0.991
b Test E-score/TOVA Reaction Time Variability/BAASFR vs. Psychometrist's Ratings	0.953	0.980
b Test E-score/TOVA Omission Errors/CAT-A Infrequency Scale – Self	0.981	0.992
b Test E-score/TOVA Omission Errors/BAASFR vs. Psychometrist's Ratings	0.959	0.983
b Test E-score/CAT-A Infrequency Scale – Self/BAASFR vs. Psychometrist's Ratings	0.946	0.977
TOVA Reaction Time Variability/CAT-A Infrequency Scale – Self/BAASFR vs. Psychometrist's Ratings	0.940	0.974
TOVA Omission Errors/CAT-A Infrequency Scale – Self/BAASFR vs. Psychometrist's Ratings	0.948	0.978

WMT = Word Memory Test, TOVA = Test of Variables of Attention, CAT-A = Clinical Assessment of Attention Deficit Disorder-Adult, BAASFR = Barkley Adult ADHD Self-report.

there is quite a significant difference in the post-test probability of suspect effort depending on the specific combination of two or three SVT measure failures. As would be expected, there is a much smaller difference between the highest and lowest post-test probabilities of malingering when failing combinations of three SVT measures.

ANOVAs indicated that there were significant differences between the ADHD credible, non-ADHD credible, ADHD suspect, and non-ADHD suspect groups on all nine of the ADHD behavior rating scale variables examined (see Table 7). The post-hoc *t*-tests show that the patients in the non-ADHD credible group repeatedly endorse experiencing significantly fewer ADHD-related symptoms than the patients in each of the other three groups on all the ADHD rating scales measures. Second, the non-ADHD suspect, ADHD credible, and ADHD suspect groups were not significantly different from one another in their level of endorsement of experiencing symptoms on seven of the eight Barkley Adult ADHD Self-report Form variables. This is also true for the nine symptoms that Barkley has recently identified as being most effective in distinguishing patients with ADHD from community control and clinical control participants. On the only Barkley Adult ADHD Self-report Form variable on which these three groups had a statistically significant difference, the ADHD suspect group reported experiencing more current (adult) inattention behaviors than the ADHD credible and non-ADD suspect groups. Thus, the non-ADHD suspect and ADHD suspect groups endorsed experiencing the same or more ADHD symptoms than the ADHD credible group in completing the ADHD behavior rating scales.

The same pattern of results was found when chi-square tests were run to determine if there were significant differences with respect to the percentage of patients within each group whose scores on the same eight Barkley Adult ADHD Self-report Form variable scores and the new Barkley nine symptom scale met the cutoff scores considered indicative of the individual having ADHD. These results show that individuals completing rating scales in an invalid manner can successfully endorse experiencing ADHD behaviors on the old and new Barkley Adult ADHD behavior rating scales such that they (a) cannot be distinguished from patients diagnosed with adult ADHD and (b) would be diagnosed with this disorder.

Many non-credible group patients in this study feigned deficits by executing tasks in a slowed manner. The non-credible group gave significantly slower performances on the Symbol Search test ($p < .001$), DKEFS-CWIT Combined Naming and Reading score ($p < .001$), and Digit Symbol test ($p < .001$). More specifically, 34% of the non-credible group gave an impaired performance (i.e., at least one standard deviation below the normative average) on Symbol Search, 40% on Digit Symbol, and 28% on DKEFS-CWIT combined naming and reading. However, the non-credible group did not exhibit significantly slower performance with respect to the CCPT hit reaction time ($p = .166$) or the NAB-NLA speed score ($p = .272$). As the data in Table 8 indicate, the Symbol Search, Digit Symbol, and DKEFS-CWIT combined naming and reading measures are not particularly sensitive in identifying suspect effort.

A major issue in research on SVT measures has been the degree to which failures on specific SVT measures are associated with poorer performance on tests in the same and unrelated cognitive domains in neuropsychological tests. Table 9

Table 7 Barkley Adult ADHD Self-report Form and new Barkley Adult ADD scale variables means and standard deviations for each diagnostic group

Group Category	1 ADD Credible	2 ADD Suspect	3 Non-ADD Credible	4 Non-ADD Suspect	Significance
Number of patients	63	12	122	35	–
Adult hyp/imp symptoms	4.06 (2.55)	5.00 (2.73)	3.39 (2.85)	5.03 (2.26)	4 > 3
Adult inattention symptoms	5.97 (2.48)	7.58 (1.56)	4.99 (3.14)	6.37 (2.25)	2 > 3
Adult hyp/imp item score	12.84 (5.66)	15.42 (6.78)	11.12 (5.82)	15.80 (5.85)	4 > 3
Adult inattention item score	16.56 (5.01)	21.67 (3.58)	14.78 (5.52)	18.29 (5.93)	2 > 1,3 & 4 > 3
Child hyp/imp symptoms	4.89 (3.28)	5.50 (2.32)	3.75 (3.08)	4.91 (3.08)	1,4 > 3
Child inattention symptoms	6.23 (3.08)	7.08 (2.02)	4.04 (2.97)	5.71 (2.76)	1,2,4 > 3
Child hyp/imp item score	14.50 (6.50)	16.58 (5.96)	11.75 (6.34)	15.26 (7.74)	1,4 > 3
Child inattention item score	16.65 (5.67)	20.83 (5.22)	12.97 (6.15)	16.43 (6.73)	1,2,4 > 3
Number of patients	50	9	81	27	–
Number of 9 new Barkley Scale symptoms endorsed	5.84 (1.77)	7.11(1.27)	4.80 (2.02)	6.63 (1.88)	1,2,4 > 3

hyp/imp = hyperactive/impulsivity.

Table 8 Sensitivity, specificity, positive and negative predictive accuracy rates for cognitive processing speed tests

Group	N	Cutoff (t-score)	Sens. (%)	Spec. (%)	15% base rate PPA (%)	15% base rate NPA (%)	30% base rate PPA (%)	30% base rate NPA (%)
Symbol Search	260	≤ 40	37.50	89.71	39.14	89.05	60.97	77.01
Digit Symbol	265	<37	25.86	93.72	42.09	87.75	63.83	74.68
DKEFT-CWIT Combined Naming and Reading	264	<44	31.03	93.69	46.46	88.50	67.82	76.02

depicts how individuals' failures on each of nine SVT measures are related to (a) their rendering statistically significant poorer performances as well as (b) their rendering impaired performances (average t-score < 40) on tests in 10 different cognitive ability domains. It is important to reiterate that the data in Table 9 are derived from ANCOVAs with Full Scale IQ as a covariate. These data reveal several interesting findings.

First, in general, individuals feigning deficits while undergoing ADHD assessment appear to feign deficits—i.e., render impaired performances—on tests that appear to directly assess attention and cognitive processing speed. They render impaired performances on tests of sustained attention, divided attention, working memory, cognitive processing speed, and memory, but not verbal comprehension/ expression, visuo-spatial ability, and executive function abilities. Second, individuals failing two of the SVT measures appear to be particularly strategic in the manner in which they feign cognitive deficits. Failure on the Word Memory Test is associated with feigning cognitive deficits only on tests of memory and sustained attention. Suspect completion of ADHD behavior rating scales as denoted by failing the CAT-A Infrequency scale is associated with feigning cognitive deficits only on tests of sustained attention. Third, individuals whose ratings of distractibility, excessive talking, and restlessness on the BAASRF scales are significantly different than the psychometrists' ratings of the same behaviors do not render poorer or impaired performances on any test in any cognitive domain.

DISCUSSION

The current study is the first of its kind to employ multiple measures to detect the exaggeration or feigning of cognitive deficits as well as invalid completion of ADHD behavior rating scales by actual patients undergoing evaluation for adult ADHD. It is also the first such study to classify such patients as making a non-credible or suspect effort in ADHD assessment based on their meeting the Slick et al. (1999) criteria for probable malingered neurocognitive dysfunction. These strengths of the study very probably improve the accuracy of the overall rate of suspect effort found in this large ($n = 268$) group of patients.

The rate of suspect effort on cognitive testing and in completing behavior rating scales in this group of patients (22%) is slightly higher than the estimated

Table 9 Number of cognitive tests in each cognitive domain in which patients failing SVT measures do significantly worse (first #) and render impaired performances (second #)

Cognitive domains (# measures in each domain)	b Test E- score	TOVA Reaction Time Variability	TOVA Omission Errors	Word Memory Test	CAT Infrequency Scale	Reliable Digit Span	Sentence Repetition	Dot Counting	BAASFR vs psychometrist's ratings
Intelligence (1)	1	1	1	0	1	1	1	1	0
	0	0	0	0	0	0	0	0	0
Verbal comprehension (3)	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0
Visuo-spatial abilities (1)	0	0	0	0	1	0	0	0	0
	0	0	0	0	0	0	0	0	0
Cognitive processing speed (5)	3	3	3	3	0	2	0	3	0
	1	0	0	0	0	0	0	0	0
Working memory (4)	3	2	2	1	0	2	0	2	0
	2	1	0	0	0	2	1	2	0
Verbal attention capacity (3)	0	1	2	1	1	1	1	2	0
	0	0	0	0	0	1	0	2	0
Memory tests (6)	5	3	3	5	0	0	0	1	0
	5	1	1	5	0	0	0	1	0
Executive function tests (7)	2	0	2	2	0	1	0	0	0
	0	0	0	0	0	0	0	0	0
Divided attention tests (3)	2	2	2	0	0	1	0	2	0
	1	1	1	0	0	1	0	2	0
Sustained attention tests (11)	6	4	5	3	5	0	0	4	1
	6	4	5	2	2	0	0	4	0

base rate of 15% malingering in general clinical patient populations not involved in litigation or criminal proceedings (Rogers et al., 1993). A total of 17% of this study's patients currently attending college gave a suspect effort in completing their ADHD evaluations. This is a far lower rate of suspect effort than those found in the two previous studies involving college students referred for ADHD assessment. Sullivan et al. (2007) and Suhr et al. (2008) classified 47% and 31% of their college students as making a suspect effort respectively.

In both the Sullivan et al. and Suhr et al. studies students were classified as making a suspect effort solely on the basis of their failing a Word Memory Test effort measure—not two or more indices of suspect effort as in the current study. If the current study had classified patients as making a suspect effort on the basis of failing one rather than two effort measures, the overall rate of suspect effort in our group of 268 patients would have increased from 22% to 51%. A suspect effort rate of 51% is far beyond the estimated base rate of 15% malingering in general clinical populations, strongly suggesting that classifying individuals undergoing ADHD evaluations as making a suspect effort on the basis of failing one effort measure results in an unacceptable number of false positives. To look at it another way, 37% of our two credible groups failed at least one SVT measure. These findings are consistent with those of Victor et al. (2009) who found that failure on any one embedded SVT measure was common in a large clinical sample of patients with mixed acute neurologic and/or psychiatric diagnoses—41% of their credible group failed at least one SVT measure.

The data obtained in this study suggest that the Word Memory Test consistency and immediate recall scores (both with 64% sensitivity), TOVA omission errors (63% sensitivity), the Conners CPT-II omission errors (56%), and TOVA reaction time variability (54%) measures have good sensitivity, the b Test e score (47%) reasonably good sensitivity, and the dot counting e score (34%) has limited sensitivity in identifying suspect test-taking effort in adults undergoing ADHD evaluations. The CAT-A infrequency scale also had good sensitivity (58%) with reasonable specificity in identifying individuals completing ADHD behavior rating scales in an invalid manner. The difference between ratings of the same three behaviors (distractibility, excessive talking, and restlessness) by the patient in completing the Barkley Adult ADHD Self-report scales and by the psychometrist in completing behavior observation analog scales was not nearly as effective in identifying invalid completion of behavior rating scales with a sensitivity of 35% and specificity of 86%.

The sensitivities and specificities of most of these effort measures in detecting suspect effort in individuals undergoing ADHD assessment compare favorably with the sensitivity and specificity of symptom validity measures used to detect suspect effort in individuals undergoing other types of neuropsychological assessment. In their meta-analysis of the effectiveness of SVT measure studies, Vickery, Berry, Inman, Harris, and Orey (2001) found an average SVT measure sensitivity of 56% and specificity of 96%.

The results of this study provide further evidence of the usefulness of employing multiple measures of effort in the diagnosis of suspect effort or malingering in yet another clinical patient population. Larrabee (2003, 2008) and Victor et al. (2009) have previously demonstrated that the failure of combinations of two SVT measures and of three SVT measures had significantly greater positive predictive power than failure on any one SVT measure in discriminating malingerers from credible patients. They did so in studies of outpatients with a head injuries as well as a wide variety of other neurological and psychiatric conditions. Some of these outpatients were either in litigation or seeking to obtain disability status. Clearly, the same benefits of aggregating results from multiple failed SVT measures in increasing diagnostic probability and reducing diagnostic errors occur in identifying suspect effort in adults undergoing neuropsychological assessment for ADHD. Even at a relatively low base rate of malingering (i.e., 15%), the post-test probabilities of an individual making a suspect effort based on the failure of two SVT measures range from 0.710 to 0.952 while failure on three SVT measures range from 0.940 to 0.993.

A major issue in research on SVT measures has been the degree to which failures on specific SVT measures are associated with feigning of deficits on cognitive tests in the same and unrelated cognitive domains. Some research (Constantinou et al. 2005; Green et al., 2001) has suggested that failure on an SVT presumably assessing one cognitive domain is associated with lower scores on neuropsychological tests in both the same and different cognitive domains. That is, poor performance on an SVT is indicative of generalized poor performance on neuropsychological tests. These findings would suggest that individuals feigning cognitive deficits do so across multiple cognitive domains in a somewhat indiscriminate manner. On the other hand, as Larrabee (2004) and Boone (2009) have pointed out, individuals feigning neuropsychological deficits might employ several different strategies and feign deficits only in cognitive domains associated with their presenting complaints.

The results of the current study suggest that individuals feigning cognitive deficits while undergoing assessment for ADHD take a somewhat more strategic approach and feign deficits on tests that appear to more directly assess attention. That is, they most often gave a suspect effort on SVT measures obviously involving sustaining attention for longer periods of time (i.e., the TOVA omission errors and reaction time variability and the b Test). They also completed two scales clearly reporting ADHD behavioral problems (the CAT-A Infrequency scale and discrepancy between the patient's and psychometrist's ratings of their degree of distractibility, excessive talking, and restlessness) in an invalid manner. On the other hand, patients making a suspect effort in an ADHD evaluation much less frequently feigned deficits on SVT measures related to the assessment of memory (the CVLT forced choice recognition measure) or the ability to sustain attention for short periods of time (the reliable digit span and sentence repetition test measures). This finding is consistent with prior research that found RAVLT Exaggeration Index (EIAVLTX) and recognition scores, WAIS II Working Memory Index, and Digit Span based effort measures were not effective in identifying suspect effort in students referred for ADHD neuropsychological assessment (Suhr et al., 2008). The exception to this pattern is the frequency with which those making a suspect effort fail the WMT. Perhaps because of the fact that one must attend to a long list of 20 word pairs twice, the WMT is perceived as a test of sustained attention by at least some individuals feigning deficits in ADHD evaluations.

Individuals feigning cognitive deficits in the current study also tend to do more poorly on tests reflecting speed of cognitive processing, i.e., the Symbol Search, Digit Symbol, and DKEFS-CWIT naming and reading measures. This finding is consistent with the study by Harrison et al. (2007) in which 31% of students reported intentionally completing tasks slowly in their attempt to simulate ADHD deficits. The reason why some individuals choose to execute tasks more slowly is unclear, however. Slowed thinking appears to be perceived as a general deficit associated with ''brain damage'' in individuals malingering traumatic brain injuries. Consequently, they often adapt a strategy of responding slowly to questions and executing timed tasks in a slowed manner (Iverson, 1995; Meyers, 2007) However, it is also conceivable that individuals feigning ADHD do tests such as Symbol Search, Digit Symbol, and the DKEFS – CWIT more slowly because they correctly perceive that difficulties in sustaining attention would, in fact, slow performance on these particular tests. Unfortunately, however, cutoff scores derived from these three measures were not very sensitive in identifying suspect effort in the current study.

The results of the current study confirm the findings of earlier research that individuals feigning or exaggerating ADHD related behavioral problems can complete adult ADHD behavior rating scales in a manner that is indistinguishable from patients with ADHD and will lead to their being diagnosed with this disorder. The Barkley Adult ADHD Self-Report Form for current symptoms (Barkley & Murphy, 2006) employed in the current study contains the same nine inattention and nine hyperactivity/impulsivity items and is scored in the same manner as the Barkley ADHD Behavior Checklist-Current Scale (Murphy & Barkley, 1996). Fisher and Watkins (2008) found that 77% of the students instructed to complete the scales in a manner to be diagnosed with ADHD did so successfully on the Barkley ADHD Behavior Checklist-Current Scale. As was the case with those patients making a suspect effort in the current study, Quinn (2003) found that students instructed to feign ADHD on the Barkley ADHD Behavior Checklist-Current Scale could not be distinguished from students diagnosed with ADHD. Harrison et al. (2007) also found that undergraduate students instructed to fake ADHD actually endorsed having more ADHD symptoms than patients previously diagnosed with ADHD on the Conners Adult ADHD Rating Scale (Conners et al., 1999).

Of the 268 patients in this study, 80 (30%) were diagnosed with adult ADHD. There is scant information on the base rate of diagnoses of adult ADHD in clinical settings. However, the contrast between the percentages of patients undergoing an evaluation of adult ADHD diagnosed with this disorder in the current study and in the one other large study of adult ADHD is striking.

Barkley et al. (2008) reported the results of a study of 243 consecutive adult outpatient referrals to the Adult ADHD Clinic in the Department of Psychiatry at the University of Massachusetts. They reported diagnosing 146 (60%) of the referred patients as having adult ADHD—exactly twice the percentage of referred patients diagnosed with this disorder in the current study (i.e., 30%).

Both studies were based on a very large group of consecutive referrals made by other medical professionals on an outpatient basis to a neuropsychology clinic affiliated with a department of psychiatry. The mean age of those diagnosed with

ADHD in the University of Massachusetts (UMASS) study is slightly older (32 vs 28), their full scale IQ score slightly lower (107 vs 112), and their education level the same as those diagnosed with ADHD in the current study. The prevalence of current or past history of depression, anxiety disorder, and substance abuse/dependence is slightly higher in the UMASS study patients. How could two studies that are similar in several respects have such different rates of diagnosis of adult ADHD?

It is possible that the much higher rate of diagnosis of ADHD in the UMASS study is due, in part, to those referred being more highly screened for ADHD by their referral sources than those referred in the current study. It is also possible that the extended clinical interview in the UMASS study was more sensitive than the clinical interview in the current study. However, it is highly doubtful that such differences between the two studies would account for such a major difference in their rates of diagnosis of adult ADHD.

The major difference between 60% and 30% of the patients being diagnosed with adult ADHD in the UMASS and the current study respectively is most likely primarily the result of two other factors. First, the diagnosis of adult ADHD in the current study is based on exhibiting some impairment in cognitive testing as well as on reported behavioral problems associated with ADHD, while the diagnosis depends only on the latter in the UMASS study. Second, the current study contains several objective measures to identify suspect effort in the assessment process, while the UMASS study relies solely on the clinician's subjective judgment to identify a suspect patient presentation. It is therefore important to reiterate that this study found that 22% of patients referred for ADHD assessment made a suspect effort using criteria meeting those of Slick et al. (1999) for probable malingering. Given that finding, it seems reasonable to conclude that at least some of the much higher rate of diagnosis of adult ADD in the UMASS study is due to patients exaggerating or feigning reported ADHD symptomatology in a clinical interview without such dishonesty being detected by the interviewer.

In conclusion, the results of the current study provide valuable information about the relative effectiveness of various embedded and free-standing SVT measures in identifying suspect effort in cognitive testing and invalid completion of self-report ADD behavior rating scales. However, the identification of such suspect effort in testing and invalid completion of behavior rating scales only assists the clinician by alerting them to the fact that the validity of the patient's cognitive test results and their completion of self-report behavior rating scales are questionable. The results of this study suggest that one can identify cognitive and behavioral symptom exaggeration in individuals undergoing an ADHD assessment with reasonable confidence. This is not to say this study's results suggest that one can classify an individual as *malingering* with reasonable confidence.

As the American Academy of Clinical Neuropsychology consensus conference statement on malingering makes very clear, the diagnosis of malingering is made very carefully not only on the basis of an individual failing SVT measures but many other factors (Heilbronner et al., 2009). Nevertheless, the results of this study suggest that a patient should not be diagnosed with ADHD if they have failed (a) two of the SVT measures or (b) one SVT measure and one measure of invalid completion of ADHD behavior ratings scales discussed in this study, unless it is the rare instance where there are several other very compelling reasons to believe that the patient has adult ADHD.

This study has several limitations. Although individuals were instructed not to take ADHD medications at the time of the assessment, there was no firm control to insure they were not taking these medications or other substances (e.g., caffeine, nicotine, illicit or prescribed drugs) that might affect performance on tests. For many patients, most/all medical, psychiatric, and educational history was provided by self-report rather than gleaned from records. It would have been helpful to have some control groups—e.g., normal participants, patients with depression, and patients with anxiety disorder—to determine whether there is some unique sensitivity and specificity of these symptom validity measures in patients referred for ADHD evaluation versus these other groups.

Patients were excluded from the study if they had histories of neurological conditions, head injuries, learning disabilities, mental retardation, a psychiatric disorder other than depression or anxiety, major substance abuse/dependence, or a physical illness at a level of severity that could have caused significant cognitive deficits, including those cognitive deficits associated with ADHD. Therefore the current findings regarding the effectiveness of SVT measures would not apply to individuals undergoing ADHD assessment who also have any of these conditions.

The primary limitation of the current study is that several of the SVT and behavior rating scales measures whose sensitivity and specificity were of interest in this study were also used to diagnose the patients with ADHD and identify cognitive and behavioral symptom exaggeration in the first place. There is the possibility that a symptom validity measure has good sensitivity simply because it was also used to identify cognitive and behavioral symptom exaggeration in the first place. The current study's design attempted to overcome this problem, at least to some degree, by diagnosing symptom exaggeration only when at least two symptom validity measures were failed. A total of 28 (48%) of the patients were classified as exaggerating cognitive and behavioral symptoms because they failed only two symptom validity measures, while 52% were diagnosed by failing at least three such measures. Of the 28 patients classified on the basis of failing only two symptom validity measures, 8 failed two SVT measures and nothing more, 8 failed one SVT measure and one cognitive test 2 *SD* below the norm for ADHD patients, 6 failed one SVT measure and had a significant discrepancy between the Barkley ADHD rating scales and the psychometrist's behavior observations, 5 failed one SVT measure and the CAT-A infrequency scale, and 1 failed one cognitive test and had a significant discrepancy between the Barkley ADHD rating scales and the psychometrist's behavior observations. The diversity of the combinations of two symptom validity measures by which individuals were classified as exaggerating cognitive and behavioral symptoms reduces somewhat the likelihood that the sensitivity rates of symptom validity measures are inflated by their being used to make the diagnosis of symptom exaggeration in the first place.

In summary, the results of the current study suggest that a significant number (i.e., 22%) of patients referred for ADHD evaluations do poorly on cognitive testing and/or complete ADHD behavior rating scales in an invalid manner in an attempt to be diagnosed with ADHD. Individuals feigning ADHD can perform on cognitive tests and complete ADHD behavior rating scales in a manner that will lead to a diagnosis of this disorder. They can easily educate themselves to be able to answer clinical interview questions in a manner generally consistent with this diagnosis. Rather than feign cognitive deficits indiscriminately, they usually give intentionally poor performances primarily on cognitive tests and SVT measures that appear to assess attention in particular. A clinician clearly needs to employ multiple embedded and free-standing SVT measures as well as an ADHD behavior rating scale validity scale to identify the exaggeration or fabrication of ADHD related symptoms.

The results of the current study strongly suggest that diagnosing ADHD on the basis of a clinical interview and ADHD behavior rating scales alone is apt to lead to an unacceptable rate of false positive diagnoses. Fortunately, there are embedded and free-standing SVT measures and an ADHD behavior rating validity scale available that have good sensitivity and specificity in identifying symptom exaggeration. It is important to employ SVT measures that appear to assess attention in particular. Given the rate of symptom exaggeration found in this large sample, it behooves clinicians to include such symptom validity measures in a comprehensive assessment of ADHD that includes a clinical interview, completion of ADHD behavior rating scales, and neuropsychological testing.

ACKNOWLEDGMENT

The authors would like to thank Steve Hughes, PhD, for his helpful comments in the editing of this manuscript.

REFERENCES

American Board of Clinical Neuropsychology (2007). American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist*, *21*, 209–231.

Arnold, G., Boone, K. B., Lu, P., Dean, A., Wen, J., Nitch, S., et al. (2005). Sensitivity and specificity of finger tapping scores for the detection of suspect effort. *The Clinical Neuropsychologist*, *19*, 105–120.

Babikian, T., Boone, K., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist*, *20*, 145–159.

Barkley, R., & Murphy, K. (2006). *Attention-deficit hyperactivity disorder: A clinical workbook – third edition*. New York: The Guilford Press.

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

Barrash, J., Suhr, J., & Manzel, K. (2004). Detecting poor effort and malingering with an expanded version of the Auditory Verbal Learning Test (AVLTX): Validation with clinical samples. *Journal of Clinical and Experimental Neuropsychology*, *26*, 125–140.

Barrett, S., Darredeau, C., Bordy, L., & Pihl, L. (2005). Characteristics of methylphenidate misuse in a university student sample. *Canadian Journal of Psychiatry, 50*, 457–461.

Boone, K. (2007) A reconsideration of the Slick et al. (1999) criteria for malingering neurocognitive dysfunction. In K. B. Boone (Ed.) *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 29–49). New York: Guilford Press.

Boone, K. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, *23*, 729–741.

Boone, K., Lu, P., & Herzberg, D. (2002a). *The Dot Counting Test*. Los Angeles, CA: Western Psychological Services.

Boone, K., Lu, P., & Herzberg, D. (2002b). *The b Test*. Los Angeles, CA: Western Psychological Services.

Boone, K., Lu, P., & Wen, J. (2005). Comparison of various RAVLT scores in the detection of noncredible memory performance. *Archives of Clinical Neuropsychology*, *20*, 301–319.

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

Brown, T. E. (1996). *Brown Attention-Deficit Disorder Scales manual*. San Antonio, TX: The Psychological Corporation.

Bush, S., Ruff, R., Troster, A., Barth, J., Koffler, S., Pliskin, N., et al. (2005). Symptom validity assessment: Practice issues and medical necessity – NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology*, *20*, 19–42.

Conners, C. (2000). *Conners Continuous Performance Test II*. Tonawanda, NY: Multi-Health Systems, Inc.

Conners, K., Erhardt, D., & Sparrow, E. (1999). *Conners Adult ADHD Rating Scales technical manual*. Tonawanda, NY: Multi-Health Systems, Inc.

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J., & McCaffrey, R. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology*, *20*, 191–198.

Conti, R. (2004). Malingered ADHD in adolescents diagnosed with conduct disorder: A brief note. *Psychological Reports*, *94*, 987–988.

Dean, A., Victor, T., Boone, K., & Arnold, G. (2008). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist*, *22*, 705–722.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System*. San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test* (2nd ed.). San Antonio, TX: The Psychological Corporation, Harcourt Assessment Company.

Diehr, M., Cherner, M., Wolfson, T., Miller, W., Grant, I., & Heaton, R. (2003). The 50 and 100 item short forms of the paced auditory serial addition task (PASAT): Demographically corrected norms and comparisons with the full PASAT in normal and clinical samples. *Journal of Clinical and Experimental Neuropsychology*, *25*(4), 571–585.

Etherton, J. L., Bianchini, K. J., Ciota, M. A., Heinly, M. T., & Greve, K. W. (2006). Pain, malingering, and the WAIS-III Working Memory Index. *Spine Journal*, *6*, 61–71.

Fisher, A., & Watkins, M. (2008). ADHD rating scales' susceptibility to faking in a college student sample. *Journal of Postsecondary Education and Disability*, *20*(2), 81–92.

Frazier, T., Demaree, H., & Youngstrom, E. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, *18*, 543–555.

Frazier, T., Frazier, A., Busch, R., Kerwood, M., & Demaree, H. (2008). Detection of simulated ADHD and reading disorder using symptom validity measures. *Archives of Clinical Neuropsychology*, *23*, 501–509.

Glutting, J., Monaghan, M., Adams, W., & Sheslow, D. (2002). Some psychometric properties of a system to measure ADHD among college students: Factor pattern, reliability, and one-year predictive validity. *Measurement and Evaluation in Counseling and Development*, *34*(4), 194–209.

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

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

Greenberg, L., Kindschi, C., Dupuy, T., & Corman, C. (1996). *Test of Variables of Attention*. Los Alamitos, CA: Universal Attention Disorders.

Greve, K., Bianchini, K., Mathias, C., Houston, R., & Couch, J. (2003). Detecting malingered performance on the Wechsler Adult Intelligence Scale: Validation of Mittenberg's approach in traumatic brain injury. *Archives of Clinical Neuropsychology*, *18*, 245–260.

Grimes, D., & Schulz, K. (2005). Epidemiology 3: Refining clinical diagnosis with likelihood ratios. *Lancet*, *365*, 1500–1505.

Harrison, A., Edwards, M., & Parker, K. (2007). Identifying students faking ADHD: Preliminary findings and strategies for detection. *Archives of Clinical Neuropsychology*, *22*, 577–588.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093–1129.

Hervey, A., Epstein, J., & Curry, J. (2004). Neuropsychology of adults with attention-deficit/hyperactivity disorder: A meta-analytic review. *Neuropsychology*, *18*, 485–503.

Iverson, G. (1995). Qualitative aspects of malingering memory deficits. *Brain Injury*, *9*(1), 35–40.

Iverson, G., & Franzen, M. (1994). The recognition memory test, digit span, and Knox cube test as markers of malingered memory impairment. *Assessment*, *1*, 323–334.

Kim, M., Boone, K., Victor, T., Marion, S., Amano, S., Cottingham, M., et al. (2010). The Warrington Recognition Memory Test for Words as a measure of response bias: Total score and response time cutoffs developed on ''real world'' credible and non-credible subjects. *Archives of Clinical Neuropsychology*, *25*, 60–70.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns of standard neuropsychological tests. *The Clinical Neuropsychologist*, *17*, 410–442.

Larrabee, G. J. (2004) *Identification of subtypes of malingered neurocognitive dysfunction*. Presented at the 32nd Annual Meeting of the International Neuropsychology Society, Baltimore, Maryland.

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

Marshall, P., & Happe, M. (2007). The performance of individuals with mental retardation on cognitive tests assessing effort and motivation. *The Clinical Neuropsychologist*, *21*, 826–840.

Meyers, J. E. (2007). Malingering mild traumatic brain injury: Behavioral approaches used by both malingering actors and probable malingerers. In K. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 239–258). New York: The Guilford Press.

Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex Figure Test and Recognition Trial*. Odessa, FL: Psychological Assessment Resources.

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

Morey, L. (1996). Detection of feigned mental disorders of the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment*, *67*, 629–640.

Murphy, K., & Barkley, R. (1996). Prevalence of DSM-IV symptoms of ADHD adult licensed drivers: Implications for clinical diagnosis. *Journal of Attention Disorders*, *1*, 147–161.

Nelson, N. W., Boone, K., Dueck, A., Wagner, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspect effort. *The Clinical Neuropsychologist*, *17*(2), 263–272.

Nigg, J. (2005). Neuropsychologic theory and findings in attention deficit/hyperactivity disorder: The state of the field and salient challenges for the coming decade. *Biological Psychiatry*, *57*, 1424–1435.

Quinn, C. (2003). Detection of malingering in assessment of adult ADHD. *The Archives of Clinical Neuropsychology*, *18*, 379–395.

Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairments: A critical review of methodological and clinical considerations. *Clinical Psychology Review*, *13*, 255–274.

Root, J., Robbins, R., Chang, L., & van Gorp, W. (2006). Detection of inadequate effort on the California Verbal Learning Test (2nd ed.): Forced choice recognition and critical item analysis. *Journal of the International Neuropsychology Society*, *12*, 688–696.

Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology*, *15*, 349–359.

Salthouse, T. (1994). The aging of working memory. *Neuropsychology*, *8*, 535–543.

Schroeder, R., & Marshall, P. (2010). Validation of the sentence repetition test as a measure of suspect effort. *The Clinical Neuropsychologist*, *24*, 326–343.

Slick, D., Sherman, E., & Iverson, G. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

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

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

Swanson, J., & Volkow, N. (2003). Serum and brain concentrations of methylphenidate: Implications for use and abuse. *Neuroscience and Biobehavioral Reviews*, *27*, 615–621.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist*, *16*, 495–505.

Tombaugh, T. N. (1996). *Test of Memory Malingering*. Los Angeles, CA: Western Psychological Services.

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

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, M. A., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, *23*, 297–313.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wetzel, L., & Boll, T. (1991). *Short Category Test, booklet format*. Los Angeles: Western Psychological Services.

White, T., & Stern, R. (2003). *Neuropsychological Assessment Battery*. Lutz, FL: Psychological Assessment Resources, Inc.

Wilens, T., Biederman, J., Faraone, S., Martelon, M., Westerberg, D., & Spencer, T. (2009). Presenting ADHD symptoms, subtypes, and co morbid disorders in clinically referred adults with ADHD. *Journal of Clinical Psychiatry*, *70*(11), 1557–1562.