Article

Development of the Subtle ADHD Malingering Screener

Assessment
2019, Vol. 26(3) 524–534

DOI: 10.1177/1073191118773881
journals.sagepub.com/home/asm
SAGE

Sujith Ramachandran[1], Erin R. Holmes[1], Meagen Rosenthal[1], Benjamin F. Banahan III[1], John Young[1], and John P. Bentley[1]

Abstract

The objective of this study was to develop a subtle self-report scale—the Subtle ADHD Malingering Screener (SAMS)—to screen for malingering among individuals reporting symptoms of attention deficit/hyperactivity disorder (ADHD). This study employed a cross-sectional experimental design with an ADHD group, a control group—comprising individuals without ADHD—and a malingering group—comprising individuals without ADHD who were instructed to feign ADHD in their responses. Factor analysis and psychometric testing were conducted to develop a final scale that could distinguish the malingering from the other groups. A 10-item, two-factor solution was obtained for the SAMS, with a sensitivity of 90.3% and specificity of 80.1%. The SAMS presents an innovative approach to help reduce overdiagnosis of ADHD and misuse of prescription stimulants. The efficient, straightforward form of the measure particularly enhances its potential application in both medical and psychosocial clinical settings.

Keywords

malingering/symptom validity testing, ADHD, drug and alcohol abuse

It is estimated that the nonmedical use of prescription stimulants among adults has increased by 67% between 2006 and 2011, and stimulant-related emergency room visits more than doubled during that same time period (Chen et al., 2016). The misuse of these medications, typically appropriately prescribed for attention deficit/hyperactivity disorder (ADHD), has been increasing on college campuses in the recent past (Chen et al., 2016; National Center on Addiction and Substance Abuse at Columbia University, 2005). Some estimates of point prevalence have indicated that such misuse can occur in up to 31% of college students (e.g., Garnier-Dykstra, Caldeira, Vincent, O'Grady, & Arria, 2012). These behaviors and trends toward higher rates over time suggest that research related to curtailing stimulant misuse is timely.

Malingering of ADHD

One approach to limiting the potential misuse of stimulant medications is identifying instances of legitimate prescription that were obtained for illicit purposes. For example, students may seek out an ADHD diagnosis to acquire prescription stimulants legally, which may be additionally incentivized by other academic accommodations such as extra time on tests (Garnier-Dykstra et al., 2012; Jasinski & Ranseen, 2011; Young & Gross, 2011; Zgierska, Miller, & Rabago, 2012). This presentation appears common in college settings, with previous research estimating that between 20% and 50% of students reporting symptoms of ADHD are either exaggerating or completely fabricating their symptoms (Harrison, 2006; Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008; Sullivan, May, & Galbally, 2007).

A more precise determination of this base rate is difficult, however, given that malingering ADHD symptoms in a medical setting can be accomplished fairly easily (Fisher & Watkins, 2008; Harp, Jasinski, Shandera-Ochsner, Mason, & Berry, 2011; Jachimowicz & Geiselman, 2004; Quinn, 2003). This has been attributed partly to taxonomic difficulty in accurately categorizing symptoms and causes of the disorder, and partly to the fact that most physicians are not adequately equipped to identify cases of malingering (Harrison, 2006; National Center on Addiction and Substance Abuse at Columbia University, 2005). The tools commonly used for diagnosing ADHD, such as self-report inventories, observer symptom ratings, and tests of certain executive functions, can all be faked easily (Fisher & Watkins, 2008; Harp et al., 2011; Jachimowicz & Geiselman, 2004; Quinn, 2003). Symptom validity tests (SVTs) often perform better at detection of malingering, but they are

[1]University of Mississippi, University, MS, USA

Corresponding Author:
Sujith Ramachandran, Department of Pharmacy Administration, School of Pharmacy, University of Mississippi, 236 Faser Hall, University, MS 38655, USA.
Email: sramacha@olemiss.edu

usually very expensive, time-consuming, and require an expert for administration and interpretation, which is not realistic in most primary care settings (Jasinski & Ranseen, 2011).

The disadvantages of SVTs can be overcome by subtle scales, which are more sensitive toward malingering because of their lack of face validity (Burkhart, Gynther, & Christian, 1978). The validity subscales from the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1940) and the Substance Abuse Subtle Screening Inventory (Miller, 1985) are examples of existing tests for the detection of malingering in reporting symptoms of mental dysfunction. In particular, the widespread use of the MMPI-2 in clinical settings has contributed to substantial research on the utility of subtle scales, which have often been tested for identification of malingering of ADHD. These examinations have typically indicated poor sensitivity (Harp et al., 2011; Rios & Morey, 2013; Young & Gross, 2011), although this should not be extrapolated to indicate that subtle scales lack value in this domain. Rather, these results are reflective of the fact that existing subtle scales for ADHD have generally not been tailored to assess a focal construct (e.g., malingering).

As such, a specific measure to detect ADHD malingering is needed to address the growing problem of stimulant misuse (Musso & Gouvier, 2012). Ideally, such a tool would be short, economical, and easy to administer, score, and evaluate to engender its use in medical and/or pharmacy settings. The primary objective of this study was to develop a self-reported malingering screener, which was named the Subtle ADHD Malingering Screener (SAMS). The secondary objective of this study was to understand the signal detection capability (i.e., sensitivity and specificity) of the SAMS compared with another commonly used measure of symptom-report validity.

Method

Study Design

This study used a cross-sectional experimental design that relied on a self-administered, computer-based survey of college students with and without ADHD. Approval for this study was obtained from the University of Mississippi Institutional Review Board.

Item Development

Potential SAMS items were developed from the theoretical perspective of the Accuracy of Knowledge framework, which indicates that malingerers can be identified by "the assessment of a person's level of relevant knowledge of the target condition that [they are] attempting to simulate" (Lanyon, 1997, p. 379). Items were generated through an extensive literature review, consultation with clinical psychological researchers and practitioners, and a series of in-depth, qualitative interviews conducted among students with and without ADHD (Haynes, Richard, & Kubany, 1995). All items were then subjected to expert review, reduced in terms of content redundancy, and refined for clarity of wording. A 7-point response format from "Always False" to "Always True" was then appended to each item in order to allow optimal variability and a true neutral point on the response scale (Comrey, 1988). Qualitative pretests were conducted using cognitive interviews to help refine the item pool (Brod, Tesler, & Christensen, 2009; Drennan, 2003; Willis, 2005).

Experimental Design

This study employed a between-subjects design with alternate instruction sets (Cofer, Chance, & Judson, 1949; Lanyon, 1970; Lees-Haley, English, & Glenn, 1991; Myerholtz & Rosenberg, 1997, 1998; Wooley, Rogers, Fiduccia, & Kelsey, 2012). Respondents were first asked to self-report whether they had any previous, formal diagnosis of ADHD. To reduce imprecision and enhance confidence in this report, respondents were also asked to clarify the approximate elapsed time since their initial ADHD diagnosis, amount of time spent during their first diagnostic appointment, and the specialty of the diagnosing practitioner. Respondents were then assigned to one of three groups as part of the experimental design. Those who self-reported previous diagnosis of ADHD were instructed to respond to the test item pool honestly, and comprised the ADHD group. Participants who reported no previous ADHD diagnosis were randomized to one of two groups where they were either (1) instructed to fake ADHD on the test item pool (malingering group; Quinn, 2003—see Supplement 1; all supplemental materials are available in the online version of the article) or (2) received generic instructions to respond to questions honestly (control group). The study's methods thus allowed differentiation of the malingering group's responses from patterns evident in people with and without ADHD, which formed the basis for iteration and refinement of the instrument.

A manipulation check was conducted using questions that were expected to be easy to malinger. The topics of these questions stemmed from the in-depth interviews conducted for the development of the item pool and included items such as "I have trouble paying attention in class." If the faking manipulation worked, the malingering group should perform similar to the ADHD group on these questions, suggesting that the instructions to malinger were clear and easy to follow. Participants in the malingering group were not given any information about symptoms of ADHD, or any other coaching, given previous results indicating that neither was necessary for successful malingering of the

condition (Booksh, Pella, Singh, & Gouvier, 2009; Rios & Morey, 2013; Tucha, Sontag, Walitza, & Lange, 2009).

Pretest

A quantitative pretest was conducted to narrow the initial 122-item pool and to filter the questions that did not differentiate well among the study groups. A total of 278 students from the University of Mississippi were recruited online and offered class credit as incentive for participation in the study. Students with ADHD were selectively recruited to ensure adequate sample size in all study groups. This resulted in 53 participants in the ADHD group, 111 in the malingering group, and 111 in the control group. The narrowing of the initial 122-item pool was conducted in line with the recommendations from Lanyon (1970). Item-level responses of the malingering and ADHD groups were compared using *t* tests to compare group means, and items with significant discrepancies were retained for further study. This process yielded a reduced pool of 38 items that were then subjected to principal components analysis (PCA) to estimate the factor structure of the scale (DeVellis, 2017). Using a combination of the latent root criterion (Hair, Black, Babin, Anderson, & Tatham, 2010; Nunnally, 1978), the scree plot (Cattell, 1966), and parallel analysis (Franklin, Gibson, Robertson, Pohlmann, & Fralish, 1995; O'Connor, 2000), a 10-item two-factor solution was retained. The first factor was composed of six items and was termed the *psychological factor* and the second factor, with four items, was labeled the *academic factor*. Both factors demonstrated satisfactory reliability with Cronbach's alpha of 0.90 and 0.87, respectively (full results provided in Supplement 2).

Participants and Study Recruitment

A new sample of participants was then recruited to further study the 10-item version of the measure. For this data collection task, study participants set up appointments with the research team to complete the survey in-person on assigned computers using the Qualtrics platform, an online survey software that enables both skip logic and random presentation of item sets. Respondents with ADHD were assigned to the ADHD group, and those without ADHD were randomly assigned to either the malingering group or the control group. Incentives in the form of class credit were provided to all respondents. Additional incentives, in the form of a chance to win one of five $25 gift cards, were offered to individuals assigned to the malingering group. Participants in this group were instructed that they would have a chance to win this incentive only if their efforts to malinger were considered successful. In reality, all participants were offered a chance to win the incentive. This deception was employed to provide tangible incentives for performance and to approximate real-world incentives for malingering.

Survey Design and Measures

In addition to the SAMS, respondents were asked to indicate their age, gender, ADHD diagnosis, and other demographic information. Additionally, to provide a comparator to the SAMS' performance in detecting malingering, the Personality Assessment Inventory (PAI; Morey, 1991) was administered at the end of the survey. This instrument was chosen because it has several validity subscales and has previously been used specifically for the evaluation of malingering of ADHD symptoms (Rios & Morey, 2013). The PAI is a 344-item scale comprising 22 subscales that is frequently used in assessment of psychopathological symptoms (e.g., anxiety, depression, mania, psychosis; Morey, 1991). Extensive psychometric research on the PAI has been conducted, which has provided strong support for its factor structure and reliability (Boone, 1998; Rogers, Ustad, & Salekin, 1998). This study used the following subscales from the PAI: Negative Impression scale (NIM), the Positive Impression Scale (PIM), the Malingering Index (MAL), and the Rogers' Discriminant Function (RDF). The PIM and the NIM capture the degree to which respondents describe themselves in an overly positive or negative light, respectively. The MAL and the RDF are supplementary validity scales that measure malingering of responses more generally (Morey, 1991). These subscales have previously shown satisfactory reliability (Morey, 1991), as well as adequate sensitivity and specificity in detecting malingering of ADHD (Rios & Morey, 2013).

Analysis

Analyses were conducted using IBM SPSS version 24.0 and Mplus version 8.0 (Muthén & Muthén, 1998-2017). Descriptive statistics were first calculated and compared across the three study groups. A CFA was then conducted in Mplus using the factor structure obtained from the PCA in the pretest and data from the ADHD and control groups—a two-factor model where each item was specified to load on just one factor. Responses from the malingering group were deliberately excluded, as this could have biased the determination of model fit (Myerholtz & Rosenberg, 1997, 1998). Finally, given the ordinal nature of the data and the presence of floor and ceiling effects on the individual items, this CFA was conducted using robust weighted least squares estimation (i.e., the WLSMV estimator in Mplus; Flora & Curran, 2004; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Model fit was estimated using the chi-square statistic, the comparative fit index (CFI), the weighted root mean square residual (WRMR; Yu, 2002), and the root mean square error of approximation (RMSEA; Hair et al., 2010). Convergent validity and discriminant validity were estimated using the Anderson and Gerbing (1988) approach. Reliability for each factor was assessed through Cronbach's alpha.

After the SAMS subscale scores were calculated, potential cutoff scores were examined with the intention of maximizing the clinical utility of the SAMS for clinical providers. A cutoff score offers a straightforward convention for a diagnosing provider to decide if the degree of malingering suggested by the SAMS warrants further action. Using the three study groups, the cutoff scores were selected to optimize both sensitivity and specificity. The receiver operator characteristic (ROC) curve, a widely used technique for this purpose, was implemented to estimate an optimal cutoff using the SAMS total scale score. Additionally, classification and regression tree (CART) analysis was conducted on each of the subscales of the SAMS. CART is a machine learning technique that provides domain-specific results that can be easily visualized and has advantages when more than two distinct groups are represented in the same data set (Weigel, Meston, & Rosen, 2005). Psychometric indices such as sensitivity (proportion of participants in the malingering group that were correctly identified by the scale as malingering), specificity (proportion of participants in the ADHD and control groups that were correctly identified by the scale as nonmalingerers), false-positive rate (proportion of participants in the ADHD and control groups that were incorrectly identified by the scale as malingering), and false-negative rate (proportion of participants in the malingering group that were incorrectly identified by the scale as nonmalingerers) were calculated based on this cutoff score. The ability of the scale to resist malingering was ultimately estimated using the false-negative rate. The degree of correspondence between the SAMS and the PAI was assessed using a phi coefficient (Cheetham & Hazel, 1969; Kuhn, 1973), and Cohen's kappa (Cohen, 1960), and the sensitivity and specificity of the two scales were compared using the McNemar test (Cheetham & Hazel, 1969).

Results

Sample Characteristics

A total of 637 respondents participated in the survey that included the 10-item SAMS and the PAI. This sample had a mean age of 20.5 years and comprised 63.6% females, 74.1% Caucasians, and 46.9% freshmen. A previous diagnosis of ADHD was reported by 16.6% ($n = 106$) of the sample. Approximately 35% of the sample self-reported misusing prescription stimulants, and eight respondents endorsed previously attempting to feign ADHD in a physician's office. These eight participants were excluded from all subsequent analyses to ensure that their practice effects did not bias the experimental manipulation. After removing respondents who had any missing values, a total of 611 complete responses were obtained on the SAMS, and 597 complete responses were obtained on the PAI. The breakdown of each of the demographic characteristics across the three study groups is provided in Table 1.

Manipulation Checks

If the instructions were followed correctly, the malingering group should respond similar to the ADHD group on these items. The malingering group was not significantly different from the ADHD group on any of these items, except for one, indicating that the instructions were generally clearly communicated and followed (Table 2).

SAMS Items

The chi-square statistic for the CFA using the WLSMV estimator was found to be 108.153 ($df = 34$; $p < .0005$), with an RMSEA of 0.078 (90% CI = 0.062-0.094), WRMR value of 0.752, and a CFI of 0.986, indicating satisfactory model fit (Hair et al., 2010; Yu, 2002). Table 3 shows the mean subscale scores in each of the study groups, both of which significantly differentiated all three study groups in post hoc comparisons. Wording of each of the items in the SAMS is presented in Supplement 3.

Psychometric Properties

Cronbach's alpha was 0.91 for the psychological subscale and 0.92 for the academic subscale, indicating satisfactory reliability. Convergent validity was demonstrated by the consistently high standardized factor loadings noted for each item (Table 4; Anderson & Gerbing, 1988). The correlation between the two scales was found to be 0.739, significantly different from 1 ($p < .0005$), indicating discriminant validity (Brown, 2014) and providing evidence that the psychological factor and the academic factor represent distinct constructs (Anderson & Gerbing, 1988; Hair et al., 2010).

Estimation of a Cutoff Score and Psychometric Indices

As seen in Figure 1, the ROC curve analysis using the total sum scale score of the SAMS was found to provide significant discrimination between the malingering group and the honest respondents, with an area under the curve of 0.901 ($SE = 0.012$; $p < .0001$). The selected cutoff from the ROC curve analysis was a total SAMS score greater than 27; that is, respondents greater than this score were classified as malingering. This cutoff score provided the best possible combination of sensitivity (89.2%) and specificity (77.8%). (See Supplement 4 for a complete list of cutoff scores obtained.)

The risk estimate for the CART analysis was 0.237 ($SE = 0.017$), indicating that 23.7% of respondents were misclassified by this technique. Respondents were classified as malingering if their psychological scale score was greater than 15 *and* the academic scale score was greater than 7 (Figure 2). Using this classification, the resultant sensitivity

Table 1. Demographics of the Final Sample.

Characteristic	ADHD (*N* = 102), *n* (%)	Malingering (*N* = 264), *n* (%)	Control (*N* = 259), *n* (%)	Statistic	*df*	*p*
Age (*M* [*SD*]; *F*)	20.9 [2.12]	20.5 [1.68]	20.5 [1.73]	$F = 2.38$	2, 625	.093
Female	51 (49.5)	178 (68.2)	173 (65.5)	$\chi^2 = 11.65$	2	.003
Ethnicity				$\chi^2 = 28.20$	12	.002
Caucasian	88 (85.4)	175 (67)	201 (75.8)			
African American	6 (5.8)	51 (19.5)	36 (13.6)			
Hispanic or Latino/a	2 (1.9)	5 (1.9)	0 (0)			
Asian or Pacific Islander	2 (1.9)	22 (8.4)	19 (7.2)			
American Indian	0 (0)	1 (0.4)	0 (0)			
Biracial or multiracial	3 (2.9)	6 (2.3)	8 (3.0)			
Other	2 (1.9)	0 (0.4)	1 (0.4)			
School year				$\chi^2 = 7.27$	6	.297
Freshman	42 (40.8)	119 (45.6)	135 (50.9)			
Sophomore	28 (27.2)	57 (21.8)	62 (23.4)			
Junior	20 (19.4)	60 (23.0)	41 (15.5)			
Senior (and above)	13 (12.6)	25 (9.6)	27 (10.2)			
Self-reported ADHD	103 (100)	0 (0)	0 (0)	$\chi^2 = 629.0$	2	<.001
Other mental illnesses	31 (30.4)	23 (8.8)	27 (10.2)	$\chi^2 = 33.27$	2	<.001
Time since ADHD diagnosis				—	—	—
<1 year ago	14 (13.9)	—	—			
1-2 years ago	13 (12.9)	—	—			
3-5 years ago	27 (26.7)	—	—			
>5 years ago	47 (46.5)	—	—			
Type of physician who diagnosed ADHD				—	—	—
Primary care provider	11 (10.8)	—	—			
Specialist	83 (81.4)	—	—			
Not sure	8 (7.8)	—	—			
ADHD diagnosis visit time				—	—	—
<30 minutes	20 (19.4)	—	—			
30-60 minutes	36 (35.0)	—	-			
60-120 minutes	29 (28.2)	—	-			
>120 minutes	10 (9.7)	—	-			
Use of stimulants with a valid Rx	96 (94.1)	9 (3.5)	4 (1.5)	$\chi^2 = 503.0$	4	<.001
Use of stimulants without a valid Rx	41 (40.2)	81 (31.2)	97 (36.6)	$\chi^2 = 4.50$	4	.342

Note. ADHD = attention deficit/hyperactivity disorder; *df* = degrees of freedom.

Table 2. Differences Between the Three Groups Among the Items Used in the Manipulation Check.

Item	ADHD, *M* (*SD*)	Malingering, *M* (*SD*)	Control, *M* (*SD*)	ADHD vs. malingering *p*	ADHD vs. control *p*	Malingering vs. control *p*
I have ADD/ADHD	6.5 (0.9)	5.2 (1.5)	1.5 (1.1)	<.01	<.01	<.01
I have difficulty keeping my focus while reading	6.2 (0.9)	5.9 (1.1)	3.8 (1.8)	.10	<.01	<.01
I have trouble sitting still	5.8 (1.5)	5.6 (1.3)	3.3 (1.8)	.61	<.01	<.01
I tend to act impulsively	5.2 (1.8)	5.1 (1.3)	2.9 (1.7)	.63	<.01	<.01
I have trouble paying attention in class	6.1 (0.9)	6.0 (1.0)	3.5 (1.8)	.65	<.01	<.01

Note. Each item was scored on a scale of 1 through 7. ADD = attention deficit disorder; ADHD = attention deficit/hyperactivity disorder.

of the measure was 90.3% and the specificity was 80.1%. In both cases, performance using subscale scores was superior to the results produced using the total scale score and the cutoff from the ROC curve analysis. Additionally, the SAMS was found to have a false-positive rate of 19.9% and a false-negative rate of 9.7% when using the cutoff score

Table 3. Mean Sum Scale Scores on Each Factor of the Subtle ADHD Malingering Screener in Each Study Group.

Group	Psychological factor, *M* (*SD*)	Academic factor, *M* (*SD*)
ADHD group	19.7 (7.0)	10.6 (5.6)
Malingering group	26.4 (7.1)	15.8 (5.7)
Control group	11.0 (5.8)	5.5 (3.0)
Total	18.9 (9.6)	10.7 (6.7)

Note. Both the psychological factor and the academic factor significantly differentiated all three study groups in post hoc comparisons ($p < .05$). The minimum and maximum possible scores for each subscale are as follows: psychological factor = 6 to 42; academic factor = 4 to 28. ADHD = attention deficit/hyperactivity disorder.

Table 4. Loadings From Confirmatory Factor Analysis as Well as Means, Standard Deviations, and Effect Sizes for Each of the Three Study Groups on Each Item in the Subtle ADHD Malingering Scale.

Item	Subscale	Standardized loadings	ADHD, *M* (*SD*)	Malingering, *M* (*SD*)	Control, *M* (*SD*)	ADHD vs. malingering, Cohen's *d*	ADHD vs. control, Cohen's *d*	Malingering vs. control, Cohen's *d*
1	PSYCH	0.772	3.9 (1.9)	5.2 (1.4)	2.1 (1.6)	−0.81	1.11	4.53
2	PSYCH	0.652	4.2 (1.9)	5.1 (1.6)	2.6 (1.9)	−0.51	0.83	4.39
3	PSYCH	0.842	3.0 (1.7)	4.4 (1.7)	1.6 (1.1)	−0.84	1.14	3.37
4	PSYCH	0.721	2.2 (1.6)	3.4 (1.7)	1.5 (1.2)	−0.67	0.53	2.57
5	PSYCH	0.759	2.8 (1.6)	4.1 (1.6)	1.6 (1.1)	−0.83	0.90	3.28
6	PSYCH	0.875	3.5 (1.6)	4.3 (1.6)	1.7 (1.3)	−0.44	1.33	3.59
7	ACAD	0.763	2.8 (2.1)	4.2 (1.7)	1.5 (1.0)	−0.73	0.97	3.18
8	ACAD	0.899	2.5 (1.6)	3.8 (1.7)	1.3 (0.8)	−0.75	1.08	2.83
9	ACAD	0.883	2.8 (1.8)	4.1 (1.7)	1.3 (0.9)	−0.79	1.15	3.10
10	ACAD	0.925	2.5 (1.6)	3.8 (1.7)	1.4 (0.9)	−0.76	0.95	2.88

Note. All items significantly differentiated all three groups in post hoc analyses. Standardized loadings are the standardized factor estimates obtained from confirmatory factor analysis. Cohen's *d* = effect size; PSYCH = Psychological subscale; ACAD = Academic subscale. Wording of the items has been provided in Supplement 3.

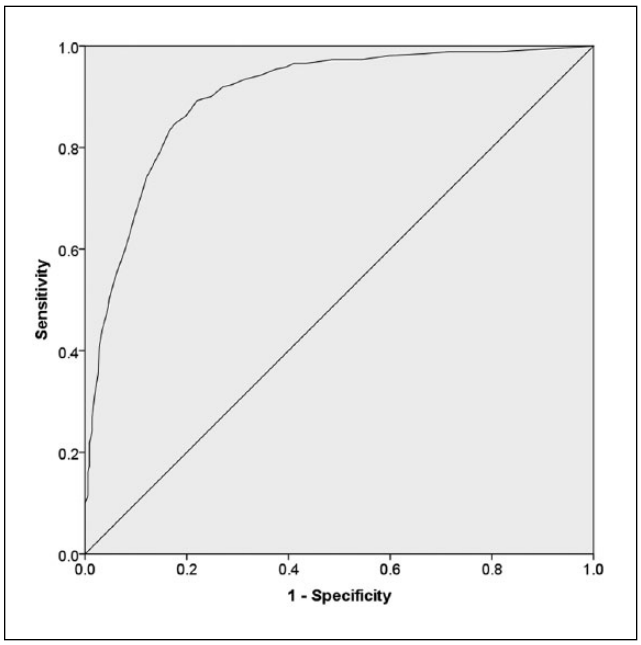

Figure 1. ROC curve analysis using SAMS total score.
Note. Area under the ROC curve = 0.901; *SE* = 0.0.012; $p < .0001$.

generated by CART analysis. Further details on the classification accuracy of the ROC and CART analyses for each study subgroup are provided in Table 5.

Comparator Scale

PAI respondents were classified as malingerers if they had an NIM score greater than 92 *or* an MAL score greater than 3. These subscale cutoffs were based on the research conducted by Rios and Morey (2013), but the rule was modified for the current study to require that only *one* of the two subscales was elevated to classify respondents' data as invalid (as opposed to *both*). This was done to provide the most liberal opportunity for the PAI validity scales to detect malingering, and thus the highest possible base rate of events to calculate signal detection statistics. Using this classification rule, the PAI was found to have a sensitivity of 51.0% and a specificity of 89.2%. Thus, the PAI's rate of false positives (10.8%) was somewhat lower than the SAMS, but the rate of false negatives (49.0%) was much higher. Using the McNemar test, the SAMS was shown to have significantly higher sensitivity ($p < .0005$) and lower specificity ($p < .0005$) than the PAI.

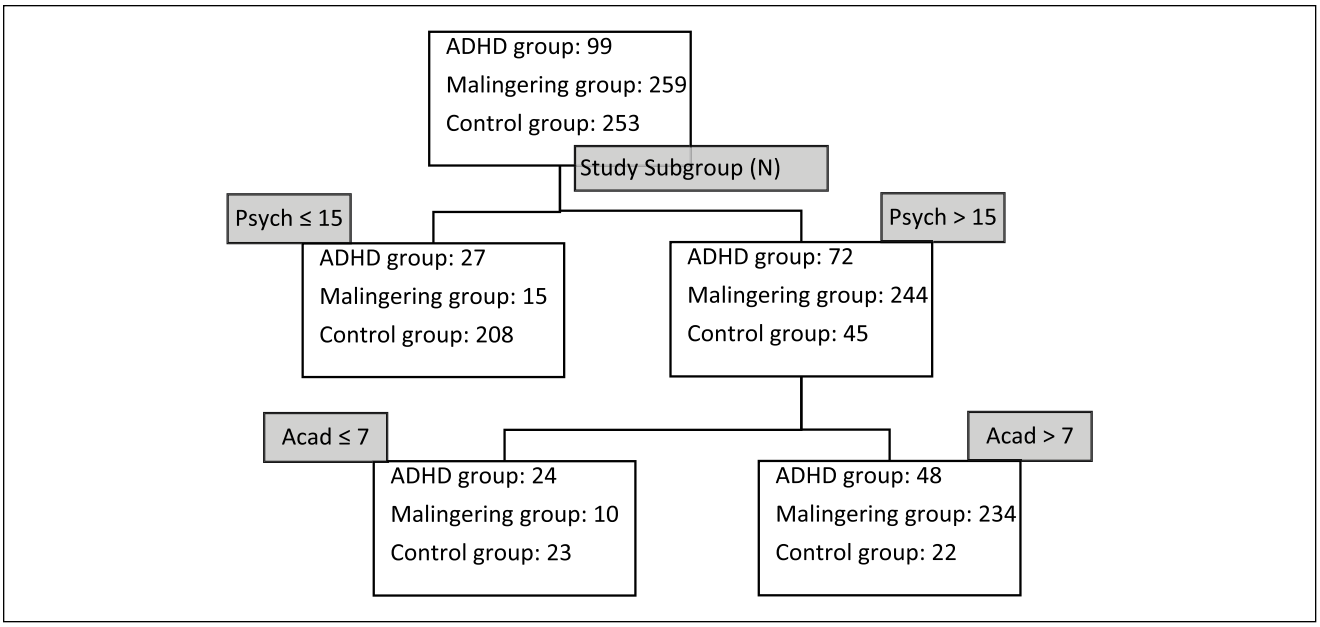

Figure 2. CART analysis for calculation of cutoff score.
Note. Psych = Psychological subscale; Acad = Academic subscale.

Finally, the total degree of agreement between the PAI and the SAMS in terms of categorical classification was found to be 70.4% (417 out of 592 respondents who had no missing responses on both the SAMS and the PAI). A phi coefficient of 0.455 and a Cohen's kappa of 0.408 were found for the two scales, indicating a moderate degree of agreement.

Discussion

This study is the first of its kind to develop a scale tailored toward identification of malingering of ADHD. The findings suggest strong psychometric properties and accurate detection of ADHD malingering in a sample that was reflective both of the university's overall population (University of Mississippi Office of Institutional Research, Effectiveness, and Planning, 2016) and the approximate base rate of stimulant misuse noted in previous research findings (i.e., 35%; Garnier-Dykstra et al., 2012).

The five items used in the manipulation check clearly indicated that the instructions were simple, clear, and easy to follow. Although the malingering and ADHD groups differed significantly (Table 2) on one manipulation check item (i.e., "I have ADD/ADHD"), this finding was not unexpected given the instructions (see Supplement 1 for the malingering instruction set). This difference did not therefore detract from the overall conclusion that the instrument, as written, would be accessible to respondents in many different settings.

The CART analysis classification rule showed moderate improvement over that of the ROC curve analysis in terms of sensitivity and specificity. It also provided an opportunity to distinguish all three study groups, instead of merely the malingering group, which could be useful to differentiate accurately diagnosed ADHD in a clinical setting. The multiple cutoff scores provided by the CART analysis can also make it harder to "coach" malingering on the SAMS, because successful malingering will require that the scores fall within a narrower range. The risk estimate from the CART analysis was found to be 0.237, indicating that 23.7% of the respondent pool was misclassified by this technique. The misclassification rate is highest in the ADHD group (48.5%; Table 5). This high false-positive rate among ADHD respondents indicates that the results of the SAMS should be interpreted with caution in the clinical setting. Being classified as a "malingerer" by the SAMS at a physician's office should signal the need for additional testing, or a referral to a specialist, in the case of primary care providers. This recommendation can maximize efficiency of resource use by only requiring additional testing or referral in some of the patients reporting ADHD symptoms, while still curtailing malingering.

Rios and Morey (2013) found that the NIM subscale in the PAI provided a sensitivity and specificity of 64% and 73%, respectively, while the MAL subscale provided a sensitivity and specificity of 38% and 81%. This study used a combination of the NIM and the MAL scales to identify malingering and found a sensitivity and specificity of 51% and 89%, respectively. The SAMS showed superiority over the PAI in identifying malingering of ADHD, demonstrating the value of a scale that is tailored to a specific condition.

Table 5. Classification Accuracy of the SAMS and PAI in Each Study Group.

Scale/Group	ADHD group, *N* (%)	Malingering group, *N* (%)	Control group, *N* (%)
ROC curve classification			
Malingerer	56 (56.6)	231 (89.2)	22 (8.7)
Honest respondent	43 (43.4)	28 (10.8)	231 (91.3)
Total	99	259	253
CART classification			
Malingerer	48 (48.5)	234 (90.3)	22 (8.7)
Honest respondent	51 (51.5)	25 (9.7)	231 (91.3)
Total	99	259	253
PAI classification			
Malingerer	14 (15.1)	129 (50.6)	23 (9.2)
Honest respondent	79 (84.9)	126 (49.4)	226 (90.8)
Total	93	255	249

Note. ROC curve classification rule: Total SAMS score greater than 27; CART classification rule: SAMS psychological factor score greater than 15 and SAMS academic factor score greater than 7; PAI classification rule: Negative Impression Scale score greater than 92 *or* Malingering scale score greater than 3. The "Total" row is the number of completed respondents for each scale, which was different for the PAI and the SAMS. SAMS = Subtle ADHD Malingering Screener; PAI = Personality Assessment Inventory; ROC = receiver operator characteristic curve; CART = classification and regression tree.

The optimal balance between sensitivity and specificity for a subtle scale depends on the purpose of the scale, particularly in terms of the recommendations for action based on its scores. Given the ethical and policy implications presented by the SAMS, it is conceived that an instrument such as the SAMS would remand respondents identified as malingering for additional testing before prescription of stimulant medications (as opposed to offering a definitive categorization as a malingerer who should not receive additional clinical consideration). Therefore, the SAMS may lead to a patient who has provided an accurate report being recommended for additional testing, which would potentially delay the onset of intervention (i.e., a false-positive result). Given the chronic, nonfatal nature of ADHD, and the intention that the SAMS is the first broadest measure in a stepped model of identification, the deleterious impact of false positives was considered to be minimal. Alternatively, a false-negative result, where a malingerer is misclassified as an honest respondent, was conceptualized as a more problematic outcome (i.e., elevating the potential for medication misuse or abuse). The instrument was thus created to be focal to the construct of malingering (as opposed to the PAI or MMPI, where validity scales are more general) and to preferentially maximize sensitivity (90.3) as opposed to specificity (80.1).

Limitations

This scale was developed using an experimental design, rather than a real-world population of people deliberately misrepresenting their symptoms. This design was chosen to facilitate ease of identification of the target malingering behavior and to help provide the required sample size (given limited likelihood of real-world malingerers presenting for study). Nonetheless, this design could have limited the representativeness of the sample, and thus the generalizability of results. For example, the demographics of the sample, while similar to the population it was drawn from, may not be representative of the population of real-world malingerers, suggesting the need for further testing in more diverse settings. Furthermore, although results from the manipulation check provide some evidence that the instructions were adequately followed, no additional efforts were made to estimate respondents' understanding of the alternate instruction set or ascertain their self-described level of effort to comply (see Supplement 1).

While time since diagnosis, specialty of diagnosing provider, and time taken for initial diagnostic visit allow for the formation of inferences about self-reported ADHD diagnosis, it might have been useful to know information about the age of onset of ADHD symptoms. It is also possible that some self-reported ADHD participants were either misdiagnosed or malingering, and some participants in the control or malingering groups could have undiagnosed ADHD. Future research involving accurate, prospective diagnostic assessment of ADHD is needed to validate the scale in groups with known diagnostic status. Finally, the trend in overdiagnosis of ADHD seen nationally may be caused by malingering, inconsistent application of diagnostic criteria, unreliable symptomatic history, misrepresentation of normative variation in attention, or a failure to recognize possible underlying substance use behavior (Angold, Erkanli, Egger, & Costello, 2000). The SAMS was designed to capture only malingering and is not likely to capture any other causes of overdiagnosis (although it may eventually be helpful as one component of a more comprehensive set of strategies to promote diagnostic accuracy).

Clinical Implications and Future Research

The efficient, straightforward form of the SAMS potentially makes it a valuable resource for primary care providers. The majority of individuals who seek assistance for a psychiatric concern tend to do so *exclusively* in this outlet (Jensen et al., 2001; Zito, Safer, Magder, Gardner, & Zarin, 1999), thus having accurate tools to aid in diagnosis and discerning necessary from illicit motivations to seek medication is critical. In this context, patients presenting with attention, hyperactivity, or impulsivity concerns who are identified by the SAMS as malingering could be recommended to undergo additional testing (either in-house in the case of providers familiar with the appropriate assessments or through referral to a specialist).

To the extent that this presentation aligns with other drug-seeking behaviors, assessed either colloquially or more formally through means of consulting state prescription monitoring programs (Perrone & Nelson, 2012), other interventions could be warranted. For example, a direct discussion of the dangers of illicit stimulant use could occur between provider and patient, or additional, informal assessment regarding substance misuse more generally could occur. Likewise, in the case of overt substance use problems being discovered, a referral to appropriate services could be made at the point of primary care contact. Future research efforts to compare the prospective results of the SAMS to a full, comprehensive assessment battery (e.g., other SVTs; the MMPI-2, self-report measures of ADHD symptoms; behavioral tasks designed to measure attention) in primary care offices will help determine its utility in these contexts.

The SAMS also presents potential for application in the pharmacy setting, where requiring administration of the screener for filling prescriptions for stimulants could present avenues to collect behavioral data on a more frequent interval and to curb drug abuse behavior. The pharmacist could also use information from the SAMS to guide patient counselling and to verify possible malingering behavior with Prescription Monitoring Program data. Finally, this study successfully demonstrates the methodology of subtle scale development. This technique can be applied in several other areas in healthcare such as the identification of opioid abusers, injection drug users, and others to implement early targeted intervention.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

The online supplementary material is available at http://journals.sagepub.com/doi/suppl/10.1177/1073191118773881.

ORCID iD

Sujith Ramachandran https://orcid.org/0000-0002-8792-4307

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411-423.

Angold, A., Erkanli, A., Egger, H. L., & Costello, J. (2000). Stimulant treatment for children: A community perspective. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 975-984.

Booksh, R. L., Pella, R. D., Singh, A. N., & Gouvier, W. D. (2009). Ability of college students to simulate ADHD on objective measures of attention. *Journal of Attention Disorders*, *31*, 325-338.

Boone, D. (1998). Internal consistency reliability of the Personality Assessment Inventory with psychiatric inpatients. *Journal of Clinical Psychology*, *54*, 839-843.

Brod, M., Tesler, L. E., & Christensen, T. L. (2009). Qualitative research and content validity: Developing best practices based on science and experience. *Quality of Life Research*, *18*, 1263-1278.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Burkhart, B. R., Gynther, M. D., & Christian, W. L. (1978). Psychological mindedness, intelligence, and item subtlety endorsement patterns on the MMPI. *Journal of Clinical Psychology*, *34*, 76-79.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Cheetham, A. H., & Hazel, J. E. (1969). Binary (presence-absence) similarity coefficients. *Journal of Paleontology*, *43*, 1130-1136.

Chen, L. Y., Crum, R. M., Strain, E. C., Alexander, G. C., Kaufmann, C., & Mojtabai, R. (2016). Prescriptions, nonmedical use, and emergency department visits involving prescription stimulants. *Journal of Clinical Psychiatry*, *77*, 297-304.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46.

Cofer, C. N., Chance, J., & Judson, A. J. (1949). A study of malingering on the Minnesota Multiphasic Personality Inventory. *Journal of Psychology*, *27*, 491-499.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, *56*, 754-761.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (Vol. 26, 4th ed.). Thousand Oaks, CA: Sage.

Drennan, J. (2003). Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing*, *42*, 57-63.

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

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466-491.

Franklin, S. B., Gibson, D. J., Robertson, P. A., Pohlmann, J. T., & Fralish, J. S. (1995). Parallel analysis: A method for determining significant principal components. *Journal of Vegetation Science*, *6*, 99-106.

Garnier-Dykstra, L. M., Caldeira, K. M., Vincent, K. B., O'Grady, K. E., & Arria, A. M. (2012). Nonmedical use of prescription stimulants during college: Four-year trends in exposure opportunity, use, motives, and sources. *Journal of American College Health*, *60*, 226-234.

Hair, J. F., Black, W.C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Pearson.

Harp, J. P., Jasinski, L. J., Shandera-Ochsner, A. L., Mason, L. H., & Berry, D. T. (2011). Detection of malingered ADHD using the MMPI-2-RF. *Psychological Injury and Law*, *4*, 32-43.

Harrison, A. G. (2006). Adults faking ADHD: You must be kidding! *The ADHD Report*, *14*(4), 1-7.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, *10*, 249-254.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*, 238-247.

Jachimowicz, G., & Geiselman, R. E. (2004). Comparison of ease of falsification of attention deficit hyperactivity disorder diagnosis using standard behavioral rating scales. *Cognitive Science Online*, *2*(1), 6-20.

Jasinski, L. J., & Ranseen, J. D. (2011). Malingered ADHD evaluations: A further complication for accommodations reviews. *The Bar Examiner*, *79*, 6-16.

Jensen, P. S., Hinshaw, S. P., Swanson, J. M., Greenhill, L. L., Conners, C. K., Arnold, L. E., . . . March, J. S. (2001). Findings from the NIMH Multimodal Treatment Study of ADHD (MTA): Implications and applications for primary care providers. *Journal of Developmental & Behavioral Pediatrics*, *22*, 60-73.

Kuhn, G. M. (1973). The phi coefficient as an index of ear differences in dichotic listening. *Cortex*, *9*, 450-457.

Lanyon, R. I. (1970). Development and validation of a psychological screening inventory. *Journal of Consulting and Clinical Psychology*, *35*, 1-24.

Lanyon, R. I. (1997). Detecting deception: Current models and directions. *Clinical Psychology: Science and Practice*, *4*, 377-387.

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

Miller, G. A. (1985). *The substance abuse subtle screening inventory (SASSI) manual*. Springville, IN: SASSI Institute.

Morey, L. C. (1991). *Personality Assessment Inventory professional manual*. Lutz, FL: Psychological Assessment Resources.

Musso, M. W., & Gouvier, W. D. (2012). Why is this so hard? A review of detection of malingered ADHD in college students. *Journal of Attention Disorders*, *18*, 186-201. doi:10.1177/1087054712441970

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Myerholtz, L. E., & Rosenberg, H. (1997). Screening DUI offenders for alcohol problems: Psychometric assessment of the substance abuse subtle screening inventory. *Psychology of Addictive Behaviors*, *11*, 155-165.

Myerholtz, L., & Rosenberg, H. (1998). Screening college students for alcohol problems: Psychometric assessment of the SASSI-2. *Journal of Studies on Alcohol and Drugs*, *59*, 439-446.

The National Center on Addiction and Substance Abuse at Columbia University. (2005). *Under the counter: The diversion and abuse of controlled prescription drugs in the U.S.* Retrieved from https://www.centeronaddiction.org/addiction-research/reports/under-counter-diversion-and-abuse-controlled-prescription-drugs-us

Nunnally, J. (1978). *Psychometric methods* (2nd ed.). New York, NY: McGraw-Hill.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*, 396-402.

Perrone, J., & Nelson, L. S. (2012). Medication reconciliation for controlled substances: An "ideal" prescription-drug monitoring program. *New England Journal of Medicine*, *366*, 2341-2343.

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

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*, 354-373.

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

Rogers, R., Ustad, K. L., & Salekin, R. T. (1998). Convergent validity of the Personality Assessment Inventory: A study of emergency referrals in a correctional setting. *Assessment*, *5*, 3-12.

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 evaluations. *Archives of Clinical Neuropsychology*, *23*, 521-530.

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

Tucha, L., Sontag, T. A., Walitza, S., & Lange, K. W. (2009). Detection of malingered attention deficit hyperactivity disorder. *ADHD Attention Deficit and Hyperactivity Disorders*, *1*, 47-53.

University of Mississippi Office of Institutional Research, Effectiveness, and Planning. (2016). *Fall 2016-2017 enrollment*. Retrieved from http://irep.olemiss.edu/fall-2016-2017-enrollment

Weigel, M., Meston, C., & Rosen, R. (2005). The Female Sexual Function Index (FSFI): Cross-validation and development of clinical cutoff scores. *Journal of Sex & Marital Therapy*, *31*, 1-20.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.

Wooley, C. N., Rogers, R., Fiduccia, C. E., & Kelsey, K. (2012). The effectiveness of substance use measures in the detection of full and partial denial of drug use. *Assessment*, *20*, 670-680. doi:10.1177/1073191112447098

Young, J. C., & Gross, A. M. (2011). Detection of response bias and noncredible performance in adult attention-deficit/hyperactivity disorder. *Archives of Clinical Neuropsychology*, *26*, 165-175.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Unpublished doctoral dissertation). University of California, Los Angeles, CA. Retrieved from https://www.statmodel.com/download/Yudissertation.pdf

Zgierska, A., Miller, M., & Rabago, D. (2012). Patient satisfaction, prescription drug abuse, and potential unintended consequences. *Journal of the American Medical Association*, *307*, 1377-1378.

Zito, J. M., Safer, D. J., Magder, L. S., Gardner, J. F., & Zarin, D. A. (1999). Psychotherapeutic medication patterns for youths with attention-deficit/hyperactivity disorder. *Archives of Pediatrics & Adolescent Medicine*, *153*, 1257-1263.