2017

CAARS-S:L INFREQUENCY INDEX VALIDATION: A PILOT COMPARISON OF PAPER AND ONLINE ASSESSMENTS

Elizabeth R. Wallace
University of Kentucky, liz.wallace@uky.edu
Author ORCID Identifier:
https://orcid.org/0000-0001-7680-1355
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.504

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Elizabeth R. Wallace, Student

Dr. David T. R. Berry, Major Professor

Dr. Mark Fillmore, Director of Graduate Studies

CAARS-S:L INFREQUENCY INDEX VALIDATION: A PILOT COMPARISON OF PAPER AND ONLINE ASSESSMENTS

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in the
College of Arts and Sciences
at the University of Kentucky

By

Elizabeth Roslyn Wallace

Lexington, Kentucky

Director: Dr. David T. R. Berry, Professor of Psychology

Lexington, Kentucky

2017

ABSTRACT OF THESIS

CAARS-S:L INFREQUENCY INDEX VALIDATION: A PILOT COMPARISON OF PAPER AND ONLINE ASSESSMENTS

One obstacle to the accurate diagnosis of ADHD in college students is malingering, although many symptom self-report measures do not contain feigning validity scales. The Infrequency Index (CII) for the Conners' Adult ADHD Rating Scale–Self-Report: Long Version (CAARS-S:L) was developed for this purpose, although further validation of the index is needed. Another topic of interest in ADHD malingering research is the increasing use of online assessments. Little is known about how ADHD is malingered in an online format, particularly on the CAARS-S:L. The current study aims to integrate these strands of research by examining the utility of the CII in detecting feigning and the effect of administration format on CAARS-S:L profiles. Data from 139 (27 diagnosed with ADHD, 46 without ADHD responding honestly, and 66 without ADHD instructed to feign) students were analyzed. Seventy-five completed the CAARS-S:L on paper, and 64 completed the assessment online. The clinical and feigning groups produced statistically similar elevations on seven of eight CAARS-S:L clinical scales. Administration format did not have a significant effect on the clinical scales or CII. The CII demonstrated 36% sensitivity and 85% specificity at the recommended cut score across administration formats. Specificity reached desirable levels at raised cut scores.

KEYWORDS: ADHD, malingering, online

Elizabeth Roslyn Wallace

December 5, 2017

CAARS-S:L INFREQUENCY INDEX VALIDATION: A PILOT COMPARISON OF PAPER AND ONLINE ASSESSMENTS

By

Elizabeth Roslyn Wallace

David T. R. Berry
Director of Thesis

Mark Fillmore
Director of Graduate Studies

December 5, 2017

TABLE OF CONTENTS

LIST OF TABLES

CAARS-S:L Infrequency Index Validation: A Pilot Comparison of Paper and Online Assessments

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterized by persistent symptoms of inattention and/or hyperactivity/impulsivity (APA, 2013). Although previously regarded as a disorder confined to childhood, it is now known that ADHD continues into adulthood for some individuals. Estimated prevalence rates of adult ADHD range from approximately 2.5% to 3.4% (Matte et al., 2015). The number of adults being treated pharmacologically for ADHD rose 250% between 2007 and 2011 (Marshall, Hoelzle, Heyerdahl, & Nelson, 2016). ADHD diagnoses specifically within the college-age population are also on the rise. The disorder affects approximately 2-8% of college students in the United States. Furthermore, of all college students receiving disability services on campuses, approximately 25% have been diagnosed with ADHD, a figure which is expected to increase (DuPaul, Weyandt, O'Dell, & Varejao, 2009).

Unfortunately for clinicians evaluating adults for ADHD, there are multiple obstacles to accurate diagnosis. One such challenge is malingering, which is defined as faking/exaggerating deficits for external benefit such as financial gain or avoidance of responsibilities (Green & Rabiner, 2012). Experts in the area suggest that malingering is more likely to occur in 'high-stakes' psychological evaluations, such as those that could lead to external benefits for the examinee (Slick, Sherman, & Iverson, 1999). ADHD evaluations can be considered 'high-stakes' in that diagnosed college students may be eligible to receive academic accommodations, such as additional testing time, access to a private testing room, and/or stimulant medication (Harp, Jasinski, Shandera-Ochsner,

Mason, & Berry, 2011). Access to controlled stimulant medication, such as Adderall or Ritalin, can be particularly appealing to college students. The effects of such medication include heightened and prolonged focus, which can be alluring for students in competitive academic environments. There is also growing evidence that these medications are sought by students for recreational use, which is associated with higher rates of alcohol and drug use and other risky behaviors (DeSantis & Hane, 2010). Furthermore, symptoms of ADHD are easily available online, making research on the disorder relatively easy for motivated students seeking a diagnosis (Williamson et al., 2014). Given these potential external gains and the availability of symptom information, malingering is a salient issue in this area. It has been estimated that 25-48% of college students feign deficits during self-referred ADHD evaluations (Sullivan, May, & Galbally, 2007). Thus, objective assessment for exaggerated and/or feigned ADHD symptoms is vital during the diagnostic process.

Although critical to assess, malingering in adult ADHD evaluations is difficult to identify accurately. For one, no consistent pattern of deficits associated with the disorder has been identified. Accordingly, a standard ADHD assessment battery has not been established (Harrison, Edwards, & Parker, 2007). Secondly, clinicians often rely on self-report measures for information on past and current symptom severity (Fuermaier et al., 2016). However, research indicates that symptoms of ADHD are easily feigned by college students on retrospective and current self-report measures, including the ADHD Rating Scale and BAARS-IV (Harrison et al., 2007; Marshall et al., 2016; Quinn, 2003). This is particularly problematic because ADHD self-report measures rarely include standard validity scales intended to identify potential feigners. Highlighting this concern

are reports such as that by Jachimowicz and Geiselman (2004), who found that 90% of students instructed to feign on a self-report ADHD measure were successful at producing profiles consistent with ADHD impairment. Though other validity measures (e.g. Test of Memory Malingering; Tombaugh, 1996) have demonstrated effectiveness in accurately detecting feigned ADHD (Sollman, Ranseen, & Berry, 2010), many vulnerable self-report measures continue to be widely-used and easily feigned. The results of multiple studies and the lack of feigning indicators solidify the urgency for strengthening self-report measures in ADHD evaluations.

One popular self-report measure used in ADHD evaluations is the Conners' Adult ADHD Rating Scales–Self Report: Long Version (CAARS–S:L). The CAARS-S:L in its original form includes eight clinical scales and one validity index (the Inconsistency Index). The Inconsistency Index (INC) assesses careless/random responding rather than overreporting or feigning. The measure's lack of a feigning validity scale has rendered it vulnerable to faked symptom reports. As previously mentioned, multiple studies have found few or no statistically significant differences on the CAARS clinical scales when comparing feigning and genuine ADHD groups (Harp et al., 2011; Jasinski et al., 2011; Sollman et al., 2010). Though the assessment manual warns that clinical scale scores greater than 80 could indicate feigning, it also states that such elevations could indicate extreme yet truthful symptomology (Conners, Erhardt, & Sparrow, 1999). Thus, the CAARS-S:L clinical scales and INC scores are inadequate for differentiating honest from feigned responses.

In order to address this concern, the CAARS-S:L Infrequency Index (CII) was created to detect potential feigning (Suhr, Buelow, & Riddle, 2011). The CII is composed

of 12 items rarely endorsed by typically-developing adults or genuine ADHD patients. Suhr et al. (2011) identified a cut score of 21 as producing 90% specificity for ADHD. When using T scores greater than 80 on the CAARS-S:L DSM-IV: Inattentive Symptoms clinical scale as the criterion for noncredible reporting, the CII identified feigners with 30% sensitivity and 100% specificity. When these analyses were repeated using the DSM-IV: Hyperactive-Impulsive clinical scale as the criterion, the CII identified feigners with 80% sensitivity and 93% specificity. When using failure on the Word Memory Test (WMT; Green, 2003), a well-validated performance validity test, as the criterion, the CII identified feigners with 24% sensitivity and 95% specificity (Suhr et al., 2011). In further validation work by Cook, Bolinger, and Suhr (2011), the CII demonstrated 52% sensitivity to feigning and 97% specificity for ADHD based on extreme elevations of the three CAARS-S:L clinical scales derived from DSM-IV ADHD criteria. However, subsequent validation using varied criteria for defining noncredible reporting has produced mixed results: Using the MMPI-2-RF validity scales, WMT, and Digit Span subtest of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV) to indicate malingering, the CII showed low sensitivity (range 13% to 36%) and moderate to high specificity (range 87% to 91.8%) (Cook et al., 2011). Similarly, in simulation studies, CII accuracy has been limited: Andresen (2012) did not find a statistically significant difference between feigning and ADHD groups on the CII. Fuermaier and colleagues (2016) found that participants with ADHD produced significantly lower scores on the CII than did a test-coached feigning group and a naïve feigning group. However, the CII did not explain a significant amount of variance in regression analyses above and beyond the measure's clinical scales. CII sensitivity was moderate (range 32% to 52%), whereas

specificity was inadequate (65%). Given the index's initial promise and subsequent mixed findings, additional validation of the CII is needed.

Another factor of interest in the assessment of ADHD is the growing popularity of online administration of clinical measures (Bhatara, Vogt, Patrick, Doniparthi, & Ellis, 2006; Steenhuis, Serra, Minderaa, & Hartman, 2009). In fact, there are several potential advantages to using online symptom reports, including saving client and clinician time and resources (Butcher, 2002; Steffen et al., 2014) and the finding that some current college students prefer this method of administration (Read, Farrow, Jaanimägi, & Ouimette, 2009). To date, few differences have been found between online and paper formats for measures of depression (Grieve & de Groot, 2011), panic (Carlbring et al., 2007), traumatic stress (Read et al., 2009), and other clinical constructs. Over 97% agreement for both paper and online measures has been reported when compared with in-person structured interviews (Steffen et al., 2014). Furthermore, online forms have been shown to demonstrate stronger reliability than their paper counterparts (Brock et al., 2015). However, some have raised concerns regarding equivalence between online and paper forms (Buchanan, 2002; Buchanan, 2003; Noyes & Garland, 2008), such as dissimilar factor structure in online forms converted from paper measures or unforeseen interactions between assessment medium and construct. Although research has emerged addressing some of these concerns (Gosling, Vazire, Srivastava, & John, 2004), the continuing discussion in the literature suggests that further evaluations of computerized assessments may be helpful.

Preliminary work on the equivalence of computerized vs. paper and pencil administration of the CAARS-S:L has begun to appear. Hirsch, Hauschild, Schmidt,

Baum, and Christiansen (2013) reported similar psychometric properties and factor structure across administration formats. However, these authors also indicated that three of the eight clinical scale scores were significantly higher for participants completing the measure online than on paper. Thus, further examination of the CAARS-S:L online form and its equivalence to the traditional paper form is needed.

The issue of detecting malingering on online assessments has received even less research attention than format equivalence. Available studies on the topic have suggested that administration format does not significantly affect malingering. In an investigation of test-takers' ability to fake good on a personality test and fake bad on a depression inventory, participants were able to dissimulate successfully regardless of administration format (Grieve & de Groot, 2011). Participants have endorsed greater acceptance of faking on online psychological tests compared to paper; however, the intention to fake did not differ between administration formats (Grieve & Elliott, 2013). Although faking on online tests has begun to be explored in the literature, there is little information on how administration format affects detection of malingering specifically on the CAARS-S:L.

The current study aims to integrate the various strands mentioned above: The comparability of online vs. paper administration of the CAARS-S:L, the accuracy of the proposed CAARS-S:L Infrequency Index (CII), and the effect of malingering vs. honest instructions on performance in an online format.

Aim 1: Examine the CAARS-S:L clinical scale scores produced by feigning and clinical ADHD groups.

Hypothesis 1: Clinical scale scores produced by feigning and clinical ADHD groups will not differ to a statistically significant degree.

Aim 2: Examine the ability of the CII to differentiate feigning and clinical ADHD groups.

Hypothesis 2: Relative to the clinical ADHD group, the feigning group will have a significantly greater number of participants obtaining a CII score of 21 or higher.

Aim 3: Examine the effect of administration format (paper v. online) on clinical scale and CII scores.

Hypothesis 3: Clinical scale and CII scores produced by the paper and online groups will not differ to a statistically significant degree.

Method

Participants

The present study included 139 undergraduate students at the University of Kentucky; of these, 27 had ADHD diagnoses and 112 did not. Participants were identified through a mass screening questionnaire administered to undergraduate students enrolled in a psychology course (i.e., the psychology subject pool). Subject pool participants received one research credit. In order to recruit additional students with ADHD, flyers advertising the study were posted in the University of Kentucky Department of Psychology building, other academic buildings, Disability Resource Center, Counseling Center, and Behavioral Health Clinic. Students with ADHD who responded to a flyer and who were not seeking research credit were compensated with $25 for participating.

The clinical ADHD group was comprised of 27 individuals with a documented ADHD diagnosis. A phone interview was administered to participants who indicated diagnoses of ADHD on the mass screening measure. This phone interview consisted of the DSM-5 structured interview for ADHD (Zimmerman, 2013), as well as questions regarding the participant's age and method of diagnosis and current ADHD medications. Participants with ADHD were excluded from data analysis if any of the following criteria were met: Diagnosis was based on a brief medical visit with their primary care physician; diagnosis was based solely on self-reported symptoms; or diagnostic criteria (five or more inattentive symptoms and/or five or more hyperactive/impulsive symptoms; several symptoms present prior to age 12; several symptoms present in two or more settings; evidence that the symptoms interfere with functioning; APA, 2013) were not met as indicated by the clinical interview. In accordance with DSM-5 ADHD criteria (APA, 2013), ADHD participants were also excluded if they reported any of the following: A diagnosis of schizophrenia or other psychotic disorder; diagnoses of dissociative or personality disorders; or current experience of substance intoxication/withdrawal. Participants reporting other conditions that may interfere with attention or concentration, such as learning/reading disabilities or significant history of head injury (defined as more than two concussions, a concussion within the past six months, or head injury more severe than a concussion) were also excluded. However, those who also reported diagnoses of anxiety or depression were included. Given the high comorbidity rate of anxiety and depression with ADHD (Murphy & Barkley, 1996), this criterion aimed to bolster external validity by representing the larger population of young adults with ADHD.

Non-clinical participants were excluded if they reported psychiatric or neurological disorders, reading/learning disabilities, or significant history of head injury. Included students were randomly assigned to either the honest (HON) or feigning (FGN) group. The honest group consisted of 46 individuals. This group was included as a manipulation check on the assessment protocol. The feigning group consisted of 66 individuals. Participants in the feigning group were offered a $25 case prize as an incentive to feign successfully.

Participants in the three testing groups (ADHD, HON, and FGN) were randomly assigned to complete the CAARS-S:L either online or on paper. The online measure is available through Multi-Health Systems Inc. Online Assessment Center. A username- and password-protected account was created for the completed assessments. Each participant in the online group completed the CAARS-S:L using a unique identification number to ensure anonymity.

Power analyses. Fuermaier et al. (2016) found an effect size of $d = 1.53$ when comparing feigners to participants with ADHD on the CII. An a-priori power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that 20 participants are needed to detect a significant effect of the CII with 80% power. F tests for fixed effects ANOVA were selected for this analysis. Grieve and de Groot (2011) found an effect size of $d = .24$ when comparing profiles from paper and online measures. An a-priori power analysis indicate that 539 participants are needed to detect a significant effect of administration format with 80% power. F tests for fixed effects ANOVA were selected for this analysis. The present study consists of 139 participants, which is above

the necessary sample size for the instruction set effect. However, it is below the necessary sample size for sufficient power for the administration format effect.

Measures

Pretest measures. The following pretest measures were administered: Subject pool mass screener; DSM-5 structured phone interview for participants reporting an ADHD diagnosis; informed consent document, which the examiner reviewed in person with each participant; Adult ADHD Self-Report Scale (ASRS-v1.1) Symptom Checklist Part A; and a brief demographics questionnaire. The mass screener was completed by students registering for psychology research participation. This questionnaire inquired about participants' histories of psychiatric diagnoses, head injuries, etc. A phone interview was subsequently administered to students who indicated ADHD diagnoses on the mass screener. The informed consent document provided information about the study and its risks and benefits. After participants signed the informed consent document, they completed the ASRS Part A. The six questions in Part A of the ASRS have been identified as being highly predictive of ADHD symptoms consistent with a clinical diagnosis (Kessler et al., 2005). This screener served as a check to ensure that nonclinical participants did not endorse a clinically significant level of ADHD symptoms (i.e. experiencing four or more symptoms Sometimes/Often). The demographics questionnaire inquired about participants' age, year in school, ethnicity, etc. All participants completed the pretest measures on paper under standard instructions.

Test battery. The test battery consisted of the **Conners' Adult ADHD Rating Scale–Self Report: Long Version (CAARS-S:L)**. This 66-item test measures current DSM-IV ADHD symptoms. The CAARS-S:L yields scores on eight clinical scales,

including Inattention/Memory Problems, Hyperactivity/Restlessness, and an overall ADHD index (Conners et al., 1999). The measure has been shown to have a sensitivity value of .82, specificity value of .87, and hit rate of .85 for ADHD (Erhardt, Epstein, Conners, Parker, & Sitarenios, 1999). Coefficient alphas for the assessment range from .86 to .92, with a median test-retest reliability value of .89 (Erhardt et al., 1999). The measure also contains the aforementioned Inconsistency Index, a standard validity scale that assesses careless/random responding. The CII is a new validity scale created to detect potential feigning on the CAARS-S:L. The authors of this index (Suhr et al., 2011) report that CII scores of 21 or greater indicate potential feigning. Instructions to complete the assessment differed by group assignment (see Procedure). Participants were randomly assigned to complete the measure either online or on paper.

Posttest measures. Following the CAARS-S:L, participants completed a posttest questionnaire. Participants were asked to reproduce their instructions and indicate their level of understanding/effort during testing on a 5-point Likert scale. Since deception was used in the feigning group, these participants then indicated on a form whether they wished to maintain or withdraw consent for the researchers to include their data in analyses. None of the participants withdrew consent. Those in the feigning group and non-credit-seeking participants completed a receipt form indicating that they received their $25. Lastly, debriefing forms explaining the purpose of the study were presented to participants. All posttest measures were administered on paper under standard instructions.

Procedure

The study utilized a simulation design with three groups – honest (HON), clinical (ADHD), and feigning (FGN). Following the pretest measures, instructions for completing the CAARS-S:L were presented to participants depending on their group. Those in the HON group were asked to complete the assessment honestly. Those in the ADHD group were told to complete the assessment honestly according to their unmedicated symptom experience. Participants in the FGN group were asked to complete the assessment as if they had ADHD and to do so without being detected by the examiner. As a monetary incentive for feigning, those in the FGN group were told that they would win $25 cash if they could take the CAARS-S:L in a way consistent with ADHD. In reality, all participants in this group received $25 upon completion of the study. FGN participants were then given a packet of ADHD reading materials adapted from Walls and colleagues (2017), which included a hypothetical scenario explaining the possible benefits of receiving academic accommodations/medication for ADHD and a description of typical ADHD symptoms available online. Following their review of the packet, participants completed an instruction check questionnaire. This questionnaire asked participants to summarize their instructions, recall ADHD characteristics, and write down strategies for faking the disorder. Participants then completed the CAARS-S:L either on paper or online.

The posttest questionnaire was subsequently administered at the end of the study procedures. This questionnaire asked participants to reproduce instructions for completing the CAARS-S:L and to indicate their perceived success at following instructions. Participants were asked to complete this questionnaire honestly. Lastly, the

debriefing form was presented. During the debriefing process, participants were thanked for their time and asked not to discuss the purpose of the study with others.

Results

Sample Description

Demographic data. A total of 220 participants recruited from the University of Kentucky subject pool and on-campus flyers completed the study. Of those participants, 81 were excluded for the following reasons: Inadequate effort (< 4 on a 5-point Likert scale) to follow instructions (n = 38); ASRS elevated in the HON and FGN groups (n = 31); greater than two concussions (n = 11); previously unindicated anxiety, depression, and/or neurological diagnoses indicated during testing (n = 6); concussion within the past six months (n = 3); previously unindicated ADHD diagnoses indicated in a nonclinical group during testing (n = 1); and, in accordance with recommendations in the assessment manual (Conners et al., 1999), omission of greater than five questions on the CAARS-S:L (n = 1). It should be noted that some participants were excluded for multiple reasons; thus, the number of participants meeting the above exclusion criteria is greater than the number of excluded participants. Of the 81 excluded participants, 38 were in the HON group, 4 were in the ADHD group, and 39 were in the FGN group.

Thus, data from 139 participants were included in analyses. Of these, 46 were in the HON group, 27 were in the ADHD group, and 66 were in the FGN group. Within each group, participants were randomly assigned to complete the CAARS-S:L on paper or online as follows: 27 (58.70%) HON paper, 19 (41.30%) HON online; 13 (48.10%) ADHD paper, 14 (51.90%) ADHD online; 35 (53.00%) FGN paper, 31 (47.00%) FGN online. Overall, the total sample was 23.00% male. The mean age of participants was

18.96 (*SD* = 1.34), and the mean number of years of education completed was 13.62 (*SD* = 0.89). Of the sample, 85.60% were right-handed, 1.40% had repeated a grade, and 27.30% reported having sustained one (19.40%) or two (7.90%) concussion(s). The ethnic breakdown was as follows: 80.60% Caucasian, 10.80% African American, 2.90% Hispanic/Latino, 2.90% Asian/Pacific Islander, 0.70% Native American, and 2.20% Other. The ethnic makeup of the sample approximates that of the University of Kentucky. Demographic characteristics are presented by instruction set in Table 3.1 and administration format in Table 3.2.

Demographic data correlations. As age, education, and number of concussions significantly differed across groups in the study, bivariate correlations were conducted to test for significant associations between these variables, CAARS-S:L clinical scale elevations, and CII raw scores. In the full sample, a small association was found between number of concussions and Hyperactivity/Restlessness T scores ($r = 0.20$, $p = 0.02$). No significant correlations were found within either the paper or online administration formats. Within the HON instruction set, no significant correlations were found. Within the ADHD group, the following medium associations were found: Age and Impulsivity/Emotional Lability T scores ($r = -0.48$, $p = 0.01$); age and ADHD Index T scores ($r = -0.48$, $p = 0.01$); age and CII raw score ($r = -0.41$, $p = 0.04$); and number of concussions and DSM-IV: Inattentive Symptoms T scores ($r = -0.43$, $p = 0.03$). Within the FGN instruction set, no significant associations were found.

ADHD diagnoses and comorbid conditions. Of the 27 participants with ADHD diagnoses included in the study, 4 (14.81%) reported being diagnosed with ADHD predominantly inattentive subtype, 7 (25.93%) with ADHD predominantly

Table 3.1 Demographic Characteristics of Participants by Instruction Set

		HON *n* = 46	ADHD *n* = 27	FGN *n* = 66	*F* *N* = 139	*p*
Male (%)		23.90	18.50	24.2	0.19	0.83
Age (years)	*M*	18.76[a]	20.15[b]	18.62[a]	16.05	0.00*
	SD	0.85	1.83	1.12		
Education (years)	*M*	13.50[a]	14.33[b]	13.41[a]	12.89	0.00*
	SD	0.72	1.14	0.72		
Repeated grade (%)		2.20	3.70	0.00	1.09	0.34
Right-handed (%)		84.80	88.90	84.80	0.14	0.87
Concussion (%)		13.00[a]	40.70[b]	31.38[ab]	4.07	0.02*
Ethnicity (%)					1.85	0.16
Caucasian		82.60	85.20	77.30		
African American		10.90	0.00	15.20		
Hispanic/Latino		2.20	7.40	1.50		
Asian/Pacific Islander		2.20	0.00	4.50		
Native American		0.00	0.00	1.50		
Other		2.20	7.40	0.00		

Note. HON = Honest; ADHD = ADHD; FGN = Feigning; *M* = Mean; *SD* = Standard Deviation.
* = $p < .05$.
[abc]Within each row, columns with different letters are statistically significantly ($p < .05$) different from each other using Tukey follow-up contrasts.

Table 3.2 Demographic Characteristics of Participants by Administration Format

		Paper *n* = 75	Online *n* = 64	*F* *N* = 139	*p*
Male (%)		20.00	26.60	0.83	0.36
Age (years)	*M*	18.76	19.20	3.84	0.052
	SD	0.90	1.70		
Education (years)	*M*	13.55	13.70	1.07	0.30
	SD	0.81	0.97		
Repeated grade (%)		0.00	3.10	2.42	0.12
Right-handed (%)		86.70	84.40	0.15	0.70
Concussion (%)		18.70	37.5	6.36	0.01*
Ethnicity (%)				0.04	0.85
Caucasian		80.00	81.30		
African American		10.70	10.90		
Hispanic/Latino		2.70	3.10		
Asian/Pacific Islander		4.00	1.60		
Native American		1.30	0.00		
Other		1.30	3.1		

Note. *M* = Mean; *SD* = Standard Deviation.
* = $p < .05$.

hyperactive/impulsive subtype, 3 (11.11%) with ADHD combined presentation, and 13 (48.15%) did not know their diagnostic subtype. The mean age of diagnosis was 13.15 years (*SD* = 5.45; range = 4.00 – 20.00). Of these participants, 9 (33.33%) reported being diagnosed by their general physician, 7 (25.93%) by a psychologist, 7 (25.93%) by a psychiatrist, 2 (7.41%) by a university behavioral health services professional, and 2 (7.41%) at the university counseling/psychological services center. Twenty-one participants (77.78%) reported taking medication for their ADHD. Ten (37.00%) reported taking Adderall, 7 (25.93%) Vyvanse, 3 (11.11%) Concerta, and 1 (3.70%) Vyvanse and Adderall. Seven (25.93%) participants in the ADHD group reported being diagnosed with other psychological conditions in addition to ADHD: Four (14.81%) participants reported anxiety and three (11.11%) reported anxiety and depression. One (3.70%) participant reported chronic migraines. Of the seven participants with ADHD reporting psychological comorbidities, five (18.52%) were currently being treated for those conditions.

CAARS-S:L clinical scales

Group differences by instruction set and administration format were evaluated for the eight CAARS-S:L clinical scales. To examine differences between the three instruction sets, a one-way analysis of variance (ANOVA) was conducted. For clinical scales with significant differences between the instruction sets, Tukey's post hoc procedure was used. On all clinical scales, HON participants produced significantly lower elevations than did the ADHD and FGN participants. On all clinical scales except Impulsivity/Emotional Lability, elevations between the ADHD and FGN groups did not significantly differ. On Impulsivity/Emotional Lability, FGN participants produced

significantly higher elevations (M = 64.85, SD = 11.13) than did participants with ADHD (M = 57.85, SD = 10.70). Such a difference indicates a medium effect size (Cohen's d = 0.64). These findings indicate that the FGN participants were generally able to produce CAARS-S:L profiles similar to those with ADHD diagnoses. Table 3.3 presents clinical scale elevations by instruction set.

As the CAARS-S:L manual indicates that extreme clinical scale T scores (> 80) may indicate feigning (Conners et al., 1999), the proportion of participants generating such scores was also examined in each instruction set. No participants in the HON group produced extreme scores on any clinical scale. Ten (37.04%) participants generated extreme scores on at least one clinical scale in the ADHD group, whereas 29 (43.94%) participants did so in the FGN group. This proportion difference was not statistically significant (p = 0.55).

Independent samples t tests were conducted in order to examine differences on the clinical scale elevations between paper and online administration formats. Cohen's d effect sizes are favored over tests of significance given the limited power for the administration format contrasts. In the HON group, no significant differences on clinical scale elevations were observed; thus, effect sizes were low. In the ADHD group, medium to large effect sizes were observed on the following scales: Inattention/Memory Problems (d = 0.69), DSM-IV: Inattentive Symptoms (d = 0.90), and DSM-IV: ADHD Symptoms Total (d = 0.59). In these instances, participants completing the measure on paper produced higher scores than those online. In the FGN group, one medium effect size (d = 0.57) was observed for the Hyperactivity/Restlessness scale, with participants completing

Table 3.3 CAARS-S:L Clinical Scale Scores by Instruction Set

	HON *M (SD)*	ADHD *M (SD)*	FGN *M (SD)*	*F*	*p*
Inatt./Mem. (T)	48.72 (7.15)[a]	67.52 (9.38)[b]	68.42 (9.12)[b]	79.24	0.00*
Hyper./Rest. (T)	47.98 (7.52)[a]	64.93 (5.74)[b]	65.18 (7.79)[b]	83.62	0.00*
Impuls./Emot. (T)	44.37 (7.96)[a]	57.85 (10.70)[b]	64.85 (11.13)[c]	55.90	0.00*
Self-Concept (T)	48.33 (8.25)[a]	54.89 (11.50)[b]	54.88 (8.34)[b]	8.16	0.00*
DSM-IV: Inatt. (T)	52.04 (8.10)[a]	76.48 (7.51)[b]	76.58 (11.04)[b]	101.75	0.00*
DSM-IV: Hyp.-Imp. (T)	45.67 (7.00)[a]	67.85 (8.67)[b]	69.95 (11.10)[b]	96.84	0.00*
Total ADHD Symp. (T)	49.04 (7.81)[a]	75.85 (6.43)[b]	77.02 (11.82)[b]	124.04	0.00*
ADHD Index (T)	45.78 (7.26)[a]	63.19 (8.11)[b]	66.00 (9.27)[b]	82.58	0.00*

Note. HON = Honest; ADHD = ADHD; FGN = Feigning; *M* = Mean; *SD* = Standard Deviation; T = T-score; Inatt./Mem. = Inattention/Memory Problems; Hyper./Rest. = Hyperactivity/Restlessness; Impuls./Emot. = Impulsivity/Emotional Lability; Self-Concept = Problems with Self-Concept; DSM-IV: Inatt. = DSM-IV: Inattentive Symptoms; DSM-IV: Hyp.-Imp. = DSM-IV: Hyperactive-Impulsive Symptoms; Total ADHD Symp. = DSM-IV: ADHD Symptoms Total; ADHD Index = ADHD Index.
* = $p < .05$.
[abc]Within each row, columns with different letters are statistically significantly ($p < .05$) different from each other using Tukey follow-up contrasts.

the assessment online producing higher scores than those on paper. Table 3.4 presents the Cohen's *d* effect sizes of the clinical scale score contrasts by administration format.

CAARS-S:L Infrequency Index

To examine CII raw score differences between the three instruction sets, a one-way ANOVA was conducted. In the case of significant differences between the instruction sets, Tukey's post hoc procedure was used. The HON group ($M = 5.15$, $SD = 4.15$) produced significantly lower raw scores on the CII than did the ADHD ($M = 14.96$, $SD = 5.48$) and FGN ($M = 17.42$, $SD = 6.84$) groups. The ADHD and FGN groups did not significantly differ on these raw scores ($p = 0.16$). Accordingly, the effect size was small ($d = 0.40$). Thus, the ADHD and FGN groups produced statistically similar raw scores on the CII. The recommended cut score of 21 (Suhr et al., 2011) was used to determine the proportion of participants elevating the CII within each instruction set. The proportion of participants in the ADHD group elevating the CII did not significantly differ from the HON group (2.20% and 14.80%, respectively; $p = 0.36$); however, the proportion of participants in the ADHD group elevating the CII significantly differed from the FGN group (14.80% and 36.40%, respectively; $p = 0.04$). Thus, more participants in the simulated than genuine ADHD group were flagged as malingering by this index. Table 3.5 presents the CII differences by instruction set.

Classification rates were further examined using CII elevations or extreme T scores as the feigning criterion. Thus, participants who either elevated the CII at the recommended cut score (CII raw > 21) or produced extreme clinical scale scores (T > 80) were flagged as potentially malingering. One participant (2.20%) in the HON group was identified using this CII/T score criterion. Eleven participants (40.70%) in the ADHD

Table 3.4 CAARS-S:L Clinical Scale Score Contrasts by Administration Format

	HON Paper v. HON Online (*d*)	ADHD Paper v. ADHD Online (*d*)	FGN Paper v. FGN Online (*d*)
Inatt./Mem. (T)	0.19	0.69	0.14
Hyper./Rest. (T)	0.32	0.05	0.57
Impuls./Emot. (T)	0.05	0.42	0.15
Self-Concept (T)	0.23	0.23	0.12
DSM-IV: Inatt. (T)	0.13	0.90	0.34
DSM-IV: Hyp.-Imp. (T)	0.13	0.12	0.45
Total ADHD Symp. (T)	0.21	0.59	0.42
ADHD Index (T)	0.11	0.35	0.10

Note: HON = Honest; ADHD = ADHD; FGN = Feigning; T = T-score; Inatt./Mem. = Inattention/Memory Problems; Hyper./Rest. = Hyperactivity/Restlessness; Impuls./Emot. = Impulsivity/Emotional Lability; Self-Concept = Problems with Self-Concept; DSM-IV: Inatt. = DSM-IV: Inattentive Symptoms; DSM-IV: Hyp.-Imp. = DSM-IV: Hyperactive-Impulsive Symptoms; Total ADHD Symp. = DSM-IV: ADHD Symptoms Total; ADHD Index = ADHD Index.

Table 3.5 CII Differences by Instruction Set

		HON	ADHD	FGN	*F*	*p*
CII Raw	*M*	5.15[a]	14.96[b]	17.42[b]	62.58	0.00*
	SD	4.15	5.48	6.84		
CII Elevated (%)		2.2[a]	14.8[a]	36.4[b]	11.39	0.00*

Note. HON = Honest; ADHD = ADHD; FGN = Feigning; *M* = Mean; *SD* = Standard Deviation; CII Raw = Infrequency Index raw score; CII Elevated = Percent of participants elevating the CII at the cut score of 21 (Suhr et al., 2011).

* = $p < .05$.

[abc]Within each row, columns with different letters are statistically significantly ($p < .05$) different from each other using Tukey follow-up contrasts.

group were identified compared to 33 participants (50.00%) in the FGN group. The proportion difference between the ADHD and FGN groups was not statistically significant ($p = 0.42$).

To examine the CII differences between the two administration formats, independent samples *t* tests were conducted. CII raw scores did not significantly differ between paper and online CAARS-S:L administrations for any instruction set group: HON $p = 0.58$; ADHD $p = 0.39$; FGN $p = 0.60$. Accordingly, these contrasts produced small Cohen's *d* effect sizes ($d = 0.17$, 0.34, and 0.13, respectively). The proportion of participants in the HON paper group elevating the CII at the standard cut score (> 21) did not differ from those in the HON online group (3.70% and 0.00%, respectively; $p = 0.41$). The proportion of participants in the ADHD paper group elevating the CII did not differ from those in the ADHD online group (15.38% and 14.29%, respectively; $p = 0.94$). The proportion of participants in the FGN paper group elevating the CII did not differ from those in the FGN online group (34.29% and 38.71%, respectively; $p = 0.71$). Thus, CII raw scores and elevations within instruction sets did not differ significantly between administration formats. Table 3.6 presents the CII differences by administration format.

CII test operating characteristics

The following metrics were calculated in order to evaluate the efficacy of the CII at discriminating between simulated and genuine ADHD: Sensitivity, or the proportion of feigners correctly identified; specificity, or the proportion of participants with genuine ADHD correctly identified; positive predictive power (PPP), or the proportion of participants who elevated the CII and were feigning; negative predictive power (NPP), or

Table 3.6 CII Differences by Administration Format

	HON Paper v. HON Online	ADHD Paper v. ADHD Online	FGN Paper v. FGN Online
CII Raw (*d*)	0.17	0.34	0.13
CII Elevated (*p*)	0.41	0.94	0.71

Note. HON = Honest; ADHD = ADHD; FGN = Feigning; *M* = Mean; *SD* = Standard Deviation; CII Raw = Infrequency Index raw score; CII Elevated = Percent of participants elevating the CII at the cut score of 21 (Suhr et al., 2011).

the proportion of participants who did not elevate the CII and were not feigning; hit rate, or the overall accuracy of the test in identifying simulated and genuine ADHD; incremental positive predictive power (IPPP), or the improvement in identifying feigners above and beyond using the base rate alone; and incremental negative predictive power (INPP), or the improvement in identifying genuine responders with ADHD above and beyond using the base rate alone.

A malingering base rate of 25% in college ADHD evaluations (Sullivan et al., 2007) and the recommended CII cut score of 21 or greater (Suhr et al., 2011) were used to produce these characteristics. Across administration formats, the CII demonstrated modest sensitivity (0.36) and acceptable specificity (0.85). On the paper CAARS-S:L, the CII demonstrated similarly modest sensitivity (0.34) and acceptable specificity (0.85). These values were slightly increased for the online format (sensitivity = 0.39; specificity = 0.86).

Specificity is generally emphasized in an ADHD evaluative setting in order to avoid false positives, i.e. labeling an examinee as feigning who is genuinely responding. Specificity of 90% or greater is considered desirable. The CII cut score was incrementally raised in order to increase the specificity, and test operating characteristics were again calculated. Increasing the cut score to 22 produced desirable specificity (0.93) for the CII in the online administration format while maintaining the same modest sensitivity (0.39). At this cut score, specificity remained acceptable in the paper format (0.85). Increasing the cut score to 23 or greater produced perfect specificity (1.00) for the CII in the paper administration format while decreasing sensitivity (0.20). Notably, specificity in the online format was desirable (0.93) at both cut scores 22 and 23. Table 3.7

Table 3.7 CII Operating Characteristics at Various Cut Scores

	Cut Score	Sn	Sp	PPP	NPP	Hit Rate	IPPP	INPP
Overall CII	≥ 21	0.36	0.85	0.86	0.35	0.51	0.61	-0.40
	≥ 22	0.32	0.89	0.88	0.35	0.48	0.63	-0.40
	≥ 23	0.26	0.96	0.71	0.38	0.46	0.46	-0.37
Paper CII	≥ 21	0.34	0.85	0.86	0.32	0.48	0.61	-0.43
	≥ 22	0.26	0.85	0.82	0.30	0.42	0.57	-0.45
	≥ 23	0.20	1.00	1.00	0.32	0.42	0.75	-0.43
Online CII	≥ 21	0.39	0.86	0.86	0.39	0.53	0.61	-0.36
	≥ 22	0.39	0.93	0.92	0.41	0.56	0.67	-0.34
	≥ 23	0.32	0.93	0.91	0.38	0.51	0.66	-0.37

Note. CII = CAARS-S:L Infrequency Index; Sn = Sensitivity; Sp = Specificity; PPP = Positive Predictive Power; NPP = Negative Predictive Power; IPPP = Incremental Positive Predictive Power; INPP = Incremental Negative Predictive Power.

presents test operating characteristics at various cut scores for the overall CII and the two administration formats.

Sensitivity and specificity were also calculated for the extreme clinical scale T scores (T > 80) and the CII/T score criteria at various CII cut scores. Extreme T scores on at least one clinical scale produced modest sensitivity and specificity (.44 and .63, respectively). The CII/T score criterion produced moderate sensitivity (.50) and inadequate specificity (.59) at the recommended CII cut score. When the CII cut score was raised to 22, sensitivity decreased (.48), and specificity remained inadequate (.63). When raised to 23, sensitivity decreased (.47), and inadequate specificity (.63) was maintained. Refer to Table 3.8 for sensitivity and specificity values of all feigning indicators evaluated in this study.

Pearson's chi-square tests of independence were conducted to examine whether the CII sensitivities and specificities significantly differed between paper and online formats when calculated at the same cut scores. At the cut score of 21, there was no significant difference between formats for sensitivity or specificity, $\chi^2(1, N = 93) = 0.79$, $p > .05$. This value indicates a small effect ($d = 0.19$). When raised to 22, there was no significant difference, $\chi^2(1, N = 93) = 3.75$, $p > .05$, indicating a small effect ($d = 0.41$). When raised to 23, there was no significant difference, $\chi^2(1, N = 93) = 3.04$, $p > .05$, indicating a small effect ($d = 0.37$). Ultimately, the sensitivity and specificity of the CII were statistically equivalent between administration formats across various cut scores.

Discussion

The accurate detection of malingered ADHD is a salient clinical issue, particularly in the college setting. Given the accommodations that may accompany the

Table 3.8. Sensitivity and Specificity of Three CAARS-S:L Feigning Indicators

	Cut Score	Sn	Sp
Extreme T	T > 80	0.44	0.63
Overall CII	CII ≥ 21	0.36	0.85
	CII ≥ 22	0.32	0.89
	CII ≥ 23	0.26	0.96
CII/T Score	CII ≥ 21	0.50	0.59
	CII ≥ 22	0.48	0.63
	CII ≥ 23	0.47	0.63

Note. Extreme T = Extreme T scores on CAARS-S:L clinical scales; CII = CAARS-S:L Infrequency Index; CII/T Score = CII elevated or T > 80 on clinical scale(s); Sn = Sensitivity; Sp = Specificity.

diagnosis, such as extra time on tests and prescription stimulant medication (Harp et al., 2011), measures frequently used in clinical evaluations should be equipped to detect feigning. Self-report symptom measures, which are frequently used in the assessment process, often do not contain malingering indices. Previous studies have demonstrated the susceptibility of these measures to faking (Marshall et al., 2016; Quinn, 2003). One such commonly-used and vulnerable test is the CAARS-S:L (Conners et al., 1999). In simulation studies, individuals asked to fake often produce CAARS-S:L clinical scale elevations statistically similar to participants with ADHD responding honestly (Harrison et al., 2007; Sollman et al., 2010). In keeping with these findings, the scores from the FGN group in this study were statistically similar to those of the ADHD group on seven of eight clinical scales. Further, the CAARS-S:L manual indicates that extreme clinical scale T scores (> 80) may indicate feigning or severe but genuine symptomology (Conners et al., 1999). In this study, this criterion as a malingering indicator produced modest sensitivity and specificity. Thus, the use of clinical scales alone to distinguish simulated from genuine responding is not supported.

In response to the demonstrated vulnerability of the CAARS-S:L to feigning, which was supported by this study, the Infrequency Index (CII) was created as the first fake bad scale for the measure (Suhr et al., 2011). Past investigations of the CII's ability to identify feigning have produced mixed results. In the total sample in this study, the CII demonstrated modest sensitivity (0.36) and adequate specificity (0.85) at the recommended cut score of 21 (Suhr et al., 2011). The CII demonstrated lower specificity than in some previous research (Cook et al., 2011; Suhr et al., 2011; Walls, Wallace, Brothers, & Berry, in press), yet the value was higher than in other previous work

(Fuermaier et al., 2016). CII specificity improved upon raising the cut score to 22 and reached desirable levels when raised to 23. The CII/T score criterion demonstrated better sensitivity than the CII alone; however, specificity for this criterion at all CII cut scores was inadequate. Ultimately, results support the use of the CII, although higher cut scores may be needed to achieve desirable specificity.

A parallel area of research is the use of online self-report measures and the identification of malingering in this administration format. Previous investigations have identified few score differences on measures of depression and other clinical constructs when administered on paper vs. online (Grieve & de Groot, 2011; Read et al., 2009). An examination of the clinical scales on the paper and online CAARS-S:L forms yielded similar T scores across the formats for five of eight scales when administered to nonclinical honestly-responding adults (Hirsch et al., 2013). Concerns remain, however, regarding the comparability and malingering detection capability for paper and online measures. Of the few studies that have investigated malingering online, most have found no differences in faking intent or activity across administration formats (Grieve & de Groot, 2011; Grieve & Elliott, 2013). This study aimed to expand the online assessment and malingering literature to ADHD and the CAARS-S:L specifically. In contrast to Hirsch et al. (2013), there were no significant clinical scale score differences between the HON paper vs. HON online groups. However, within the ADHD group, participants completing the paper assessment produced significantly higher elevations on the Inattention/Memory Problems, DSM-IV: Inattentive Symptoms, and DSM-IV: ADHD Symptoms Total scales than those completing the online version. These differences indicated medium to large effects. Within the FGN group, participants completing the

paper assessment produced lower scores on the Hyperactivity/Restlessness scale than those completing the online version. Such a difference indicated a medium effect. This latter finding could reflect the concern that online responders may exaggerate or disclose more severe symptoms given the perception of anonymity (Buchanan, 2002). However, the medium to large effects seen in the ADHD group do not support this concern. Future research is warranted to test the replicability of these format differences.

Regarding the malingering detection accuracy of paper vs. online forms of the CAARS-S:L, both versions produced modest sensitivity (.34 and .39, respectively) and acceptable specificity (.85 and .86, respectively) at the standard cut score of 21. Using a raised cut score of 22, the online form reached desirable specificity (.93), whereas the paper form maintained acceptable specificity (.85). Using a raised cut score of 23, the paper form reached perfect specificity (1.00). Chi-square tests of independence indicated that sensitivity and specificity values were not statistically significantly different between paper and online formats at the three applied cut scores. Thus, malingering detection accuracy did not significantly differ between versions. Ultimately, across administration formats, within-instruction set clinical scale elevations were generally similar, and the CII detected feigning at statistically equivalent levels of accuracy. These finding add to previous studies indicating support for the use of online assessments.

Additional results of interest included the demographic differences between instruction sets. The ADHD group reported a significantly greater number of concussions than the HON group, which may drive the small positive association between number of concussions and Hyperactivity/Restlessness T scores. Within the ADHD group, a medium negative association was observed between number of concussions and DSM-

IV: Inattentive Symptoms T scores. Also within the ADHD group, negative associations existed between age and Impulsivity/Emotional Lability T scores, ADHD Index T scores, and CII raw scores. These data suggest that symptoms in diagnosed individuals may lessen with age and that a relationship exists between concussions and ADHD or hyperactivity symptomology. Research indicates that ADHD symptomology does decrease with age, particularly for males presenting primarily with hyperactivity. It has also been suggested that symptoms may persist, but the diagnostic criteria for the disorder are inappropriate for and insensitive to the adult presentation (Simon, Czobar, Bálint, Mészáros, & Bitter, 2009). Further, the CAARS-S:L DSM scales were developed using DSM-IV criteria for ADHD; these associations may be different if the assessment were updated to reflect the most recent criteria. Continued research is warranted on the trajectory of ADHD symptoms through life and the optimal diagnostic criteria for identifying cases that persist into adulthood.

The nature of the concussion-ADHD relationship has begun to be explored in the literature: Research indicates that children who have sustained concussions exhibit greater inattention and impulsivity when compared to their peers and their own pre-injury behavior (McKinlay, 2014; Moore et al., 2016). Further, high school and college student athletes with ADHD are significantly more likely to report histories of concussions relative to student athletes without the disorder (Alosco, Fedor, & Gunstad, 2014; Iverson, Atkins, Zafonte, & Berkner, 2016). In a meta-analytic review, student athletes ages 12 to 25 with a history of concussion had ADHD more often than athletes reporting no history of concussion. In those with ADHD and concussion, the onset of the disorder preceded the injury, indicating that ADHD may act as a risk factor for concussion

(Biederman et al., 2015). The majority of the literature has explored the relationship between concussion and ADHD diagnoses and symptoms in student athlete populations; thus the results may not generalize to non-athletes. The potential effects of concussion history on inattention and hyperactivity/impulsivity symptoms warrant further investigation.

Strengths and Limitations

This study included the following efforts to strengthen internal validity: DSM-5 structured interview for ADHD; ASRS-v1.1 Part A symptom checklist as a screening measure for HON and FGN groups; instruction check and effort measure; and monetary incentive to feign. Efforts were also made to strengthen external validity, such as including participants with ADHD who endorsed comorbid psychological conditions and using symptom information that is available online in the FGN instruction packet. Although simulation designs allow for strong internal validity, external validity is inherently limited (Rogers & Cruise, 1998). Thus, quality known-groups design studies are needed in this area. Further, as noted previously, this study was underpowered for the administration format contrasts. Since this was a pilot study, further research is needed on the comparability of paper and online CAARS-S:L forms.

Conclusions

This study aimed to provide further validation of the CII as a malingering indicator for the CAARS-S:L and evaluate the comparability of the paper and online forms of the assessment. Results indicate that students instructed to feign ADHD were generally able to produce clinical scale scores similar to those who have been diagnosed with ADHD on paper and online formats. The CII distinguished dissimulated from

genuine ADHD with modest sensitivity and adequate specificity at the recommended cut score and performed more favorably than the extreme T score and CII/T score criteria. Specificity of the CII improved at raised cut scores. The CII performed similarly across administration formats. The results support the use of the CII to identify malingering on the CAARS-S:L, as well as the use of the online form of the assessment.

References

Alosco, M. L., Fedor, A. F., & Gunstad, J. (2014). Attention deficit hyperactivity disorder as a risk factor for concussions in NCAA division-I athletes. *Brain Injury, 28*(4), 472-474.

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

Andresen, E. N. (2012). *Effort and incentive in the assessment of symptoms of attention deficit/hyperactivity disorder*. The University of Wisconsin-Milwaukee.

Bhatara, V. S., Vogt, H. B., Patrick, S., Doniparthi, L., & Ellis, R. (2006). Acceptability of a web-based attention-deficit/hyperactivity disorder scale (T-SKAMP) by teachers: A pilot study. *The Journal of the American Board of Family Medicine*, *19*(2), 195-200.

Biederman, J., Feinberg, L., Chan, J., Adeyemo, B. O., Woodworth, K. Y., Panis, W., ... & Kenworthy, T. (2015). Mild traumatic brain injury and attention-deficit hyperactivity disorder in young student athletes. *The Journal of Nervous and Mental Disease*, *203*(11), 813.

Brock, R. L., Barry, R. A., Lawrence, E., Rolffs, J., Cerretani, J., & Zarling, A. (2015). Online administration of questionnaires assessing psychological, physical, and sexual aggression: Establishing psychometric equivalence. *Psychology of Violence, 5*(3), 294-304.

Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice*, *33*(2), 148.

Buchanan, T. (2003). Internet-based questionnaire assessment: Appropriate use in clinical contexts. *Cognitive Behaviour Therapy*, *32*(3), 100-109.

Butcher, J. N. (2002). How to use computer-based reports. *Clinical Personality Assessment*, 109-125.

Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Öst, L. G., & Andersson, G. (2007). Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior*, *23*(3), 1421 -1434.

Conners, C. K., Erhardt, D., & Sparrow, E. (1999). *Conner's adult ADHD rating scales: Technical manual*. Multi-Health Systems Incorporated (MHS).

Cook, C. M., Bolinger, E., & Suhr, J. (2016). Further validation of the Conner's Adult Attention Deficit/Hyperactivity Rating Scale Infrequency Index (CII) for detection of non-credible report of attention deficit/hyperactivity disorder symptoms. *Archives of Clinical Neuropsychology*, *31*(4), 358-364.

DeSantis, A. D., & Hane, A. C. (2010). "Adderall is definitely not a drug": Justifications for the illegal use of ADHD stimulants. *Substance Use & Misuse, 45*(1-2), 31-46.

DuPaul, G. J., Weyandt, L. L., O'Dell, S. M., & Varejao, M. (2009). College students with ADHD: Current status and future directions. *Journal of Attention Disorders*, *13*(3), 234-250.

Erhardt, D., Epstein, J. N., Conners, C. K., Parker, J., & Sitarenios, G. (1999). Self -ratings of ADHD symptoms in adults, 2: reliability, validity, and diagnostic sensitivity. *Journal of Attention Disorders*, *3*(3), 153-158.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Fuermaier, A. B., Tucha, L., Koerts, J., Weisbrod, M., Grabemann, M., Zimmermann, M., ... & Tucha, O. (2016). Evaluation of the CAARS Infrequency Index for the Detection of Noncredible ADHD Symptom Report in Adulthood. *Journal of Psychoeducational Assessment*, 1-12.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web -based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*(2), 93.

Green, P. (2003). Green's word memory test for windows: User's manual. *Edmonton, Canada: Green's Publishing*.

Green, A. L., & Rabiner, D. L. (2012). What do we really know about ADHD in college students? *Neurotherapeutics*, *9*(3), 559-568.

Grieve, R., & de Groot, H. T. (2011). Does online psychological test administration facilitate faking? *Computers in Human Behavior*, *27*(6), 2386-2391.

Grieve, R., & Elliott, J. (2013). Cyberfaking: I can, so I will? Intentions to fake in online psychological testing. *Cyberpsychology, Behavior, and Social Networking*, *16*(5), 364-369.

Harp, J. P., Jasinski, L. J., Shandera-Ochsner, A. L., Mason, L. H., & Berry, D. T. R. (2011). Detection of malingered ADHD using the MMPI-2-RF. *Psychological Injury and Law*, *4*(1), 32-43.

Harrison, A. G., Edwards, M. J., & Parker, K. C. (2007). Identifying students faking ADHD: Preliminary findings and strategies for detection. *Archives of Clinical Neuropsychology, 22*(5), 577-588.

Hirsch, O., Hauschild, F., Schmidt, M. H., Baum, E., & Christiansen, H. (2013). Comparison ofWeb-based and paper-based administration of ADHD questionnaires for adults. *Journal of Medical Internet Research*, *15*(3), e47.

Iverson, G. L., Atkins, J. E., Zafonte, R., & Berkner, P. D. (2016). Concussion history in adolescent athletes with attention-deficit hyperactivity disorder. *Journal of Neurotrauma, 33*(23), 2077-2080.

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., Harp, J. P., Berry, D. T. R., Shandera-Ochsner, A. L., Mason, L. H., & Ranseen, J. D. (2011). Using symptom validity tests to detect malingered ADHD in college students. *The Clinical Neuropsychologist*, *25*(8), 1415-1428.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E. V. A., ... & Ustun, T. B. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine, 35*(2), 245-256.

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

McKinlay, A. (2014). Long-term outcomes of traumatic brain injury in early childhood. *Australian Psychologist, 49*(6), 323-327.

Moore, D. R., Pindus, D. M., Raine, L. B., Drollette, E. S., Scudder, M. R., Ellemberg, D., & Hillman, C. H. (2016). The persistent influence of concussion on attention, executive control and neuroelectric function in preadolescent children. *International Journal of Psychophysiology, 99*, 85-95.

Murphy, K., & Barkley, R. A. (1996). Attention Deficit Hyperactivity Disorder adults: Comorbidities and adaptive impairments. *Comprehensive psychiatry, 37*(6), 393-401.

Noyes, J. M., & Garland, K. J. (2008). Computer-vs. paper-based tasks: Are they equivalent? *Ergonomics*, *51*(9), 1352-1375.

Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology*, *18*(4), 379-395.

Read, J. P., Farrow, S. M., Jaanimägi, U., & Ouimette, P. (2009). Assessing trauma and traumatic stress via the Internet: Measurement equivalence and participant reactions. *Traumatology*, *15*(1), 94-102.

Rogers, R., & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior, 22*(3), 273-285.

Simon, V., Czobor, P., Bálint, S., Mészáros, Á., & Bitter, I. (2009). Prevalence and correlates of adult attention-deficit hyperactivity disorder: Meta-analysis. *The British Journal of Psychiatry, 194*(3), 204-211.

Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*(4), 545-561.

Sollman, M. J., Ranseen, J. D., & Berry, D. T. R. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, *22*(2), 325-335.

Steenhuis, M. P., Serra, M., Minderaa, R. B., & Hartman, C. A. (2009). An internet version of the Diagnostic Interview Schedule for Children (DISC-IV): Correspondence of the ADHD section with the paper-and-pencil version. *Psychological Assessment*, *21*(2), 231-234.

Steffen, M., Murad, M. H., Hays, J. T., Newcomb, R. D., Molella, R. G., Cha, S. S., & Hagen, P. T. (2014). Self-report of tobacco use status. Population Health Management, *17*(3), 185-189.

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

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

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

Walls, B.D., Wallace, E.R., Brothers, S.L., & Berry, D.T.R. (*in press*). Utility of the CAARS Validity Scales in Identifying Simulated ADHD and Random Responding. *Psychological Assessment*.

Williamson, K. D., Combs, H. L., Berry, D. T.R., Harp, J. P., Mason, L. H., & Edmundson, M. (2014). Discriminating among ADHD alone, ADHD with a comorbid psychological disorder, and feigned ADHD in a college sample. *The Clinical Neuropsychologist, 28*(7), 1182-1196.

Zimmerman, M. (2013). Interview Guide for Evaluating DSM-5 Psychiatric Disorders and the Mental Status Examination. East Greenwich, RI: Psych Products Press.

Vita

1. Birmingham-Southern College (B.S., Psychology)
2. Jesse G. Harris Psychological Services Center: Student Therapist; Kentucky Neuroscience Institute Practicum Student; Grayson & Associates Psychological Testing Assistant
3. University of Kentucky Department of Psychology Daniel R. Reedy Quality Achievement Fellowship; Birmingham-Southern College Kurt Lewin/Richard McCallum Award in Psychology
4. Berry, D. T. R., Walls, B. D., Bouquet, C. M. & **Wallace, E. R.** (2016). Malingered Neurocognitive Deficits in Mild Traumatic Brain Injury. In D.Y Han (Ed.) Acquired Brain Injury: Clinical Essentials for Neurotrauma and Rehabilitation Professionals. New York: Springer Publishing Company.
   Walls, B. D., **Wallace, E. R.,** Brothers, S. L., Berry, D. T. R. (in press). Utility of the CAARS validity scales in identifying simulated ADHD and random responding. *Psychological Assessment*.

Elizabeth Roslyn Wallace