Item Response Theory Analyses Of Barkley's Adult ADHD Rating Scales

Morgan Nitta
Marquette University

ITEM RESPONSE THEORY ANALYSES OF BARKLEY'S ADULT ADHD RATING SCALES

by

Morgan E. Nitta

A Thesis submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin

December 2018

ABSTRACT
ITEM RESPONSE THEORY ANALYSES
OF BARKLEY'S ADULT ADHD
RATING SCALES

Morgan E. Nitta, B.S.

Marquette University, 2018

There are many challenges associated with assessment and diagnosis of ADHD in adulthood. A significant percentage of adult patients may fabricate or exaggerate ADHD symptoms when completing self-report measures in hopes of securing a diagnosis. Further, there are conflicting findings surrounding the similarity between ADHD presentation in adults and children, reflected in rating-scales and symptoms outlined in the diagnostic criteria.

This research provides novel information regarding relationships between common adult ADHD self-report form items and corresponding theoretical constructs of inattention (IA) and hyperactivity/impulsivity (H/I). Utilizing the graded response model (GRM) from item response theory (IRT), a comprehensive item-level analysis of adult ADHD rating scales in a clinical population was conducted with Barkley's Adult ADHD Rating Scale-IV, Self-Report of Current Symptoms (CSS), a self-report diagnostic checklist. A similar self-report measure quantifying retrospective report of childhood symptoms, Barkley's Adult ADHD Rating Scale-IV, Self-Report of Childhood Symptoms (BAARS-C), was also evaluated to further understand ADHD item functioning through the lifespan. Differences in item functioning were also considered after identifying and excluding individuals with suspect effort.

Results reveal that items associated with symptoms of IA and H/I are endorsed differently across the lifespan, and these data suggest that they vary in their relationship to the theoretical constructs of IA and H/I. Screening for sufficient effort did not meaningfully change item level functioning. The application IRT to direct item-to-symptom measures allows for a unique psychometric assessment of how the current *DSM-5* symptoms represent latent traits of inattention and hyperactivity/impulsivity. Meeting a symptom threshold of five or more symptoms may be misleading. Closer attention given to specific symptoms in the context of the clinical interview and reported difficulties across domains may lead to more informed diagnosis.

ACKNOWLEDGEMENTS

Morgan E. Nitta, B.S

In no particular order, I would like to thank my partner, Nicholas Kirrane for the love and support he provided for this academic achievement. Additionally, I would like to thank my parents, Kathleen and Darryl Nitta for instilling a love of learning into my life. My gratitude extends to the members of my cohort who have walked this academic path beside me, as well as the Hoelzle research lab at Marquette University. Finally, I would like to thank my thesis committee, especially my advisor, Dr. James Hoelzle for guidance, mentorship, and support.

TABLE OF CONTENTS

LIST OF TABLES

Introduction

Attention-deficit/hyperactivity disorder (ADHD; American Psychiatric Association [APA], 2013) is defined by symptoms of hyperactivity, impulsivity, and/or inattention that negatively impact functioning. Historically considered a neurodevelopmental disorder, there was a widespread belief that as children matured, the pervasiveness of symptoms would decrease or disappear (Ross & Ross, 1976). However, it is increasingly evident that ADHD persists in adulthood with prevalence rates of adult ADHD ranging from 1% to 5% (e.g., see Faraone & Biederman, 2005; Kessler et al., 2006; Kooij et al., 2005; Simon, Czobor, Bálint, Mészáros, & Bitter, 2009).

While standard diagnostic practices have been established for children with ADHD (e.g., see *Pediatrics,* 2011), a consensus statement has failed to emerge describing how to optimally and reliably evaluate adults referred for ADHD. Guidelines for diagnosis of ADHD in adults include a thorough clinical interview and the use of behavior rating scales (i.e., a diagnostic criteria checklist; Haavik, Halmoy, Lundervold, & Fasmer, 2010; Post & Kurlansik, 2012). The most frequently administered behavior rating scales ask the referred patient to indicate the presence of current ADHD symptoms and to retrospectively recall ADHD symptoms experienced prior to age 12 (e.g., Barkley Adult ADHD Rating Scales [BAARS], 2011; Wendar Utah Rating Scale [WURS], Ward, Wender, & Reimherr, 1993).

The current research focuses on exploring the psychometric properties of current and retrospective childhood self-report ADHD symptom scales in the context of two specific challenges to diagnosing adult ADHD. The first primary challenge to consider in

conducting psychometric studies is associated with valid symptom reporting. In addition to retrospective childhood symptom reports not necessarily being reliable (Mannuzza et al., 2002) and a tendency for adults to have limited insight into recognizing and quantifying inattentive symptoms (Kooij et al., 2008), there is increasing awareness of the possibility that patients may engage in symptom exaggeration during an adult ADHD evaluation (Suhr & Berry, 2017). A comprehensive literature review documents rates of empirically derived non-credible presentation ranging from approximately 8% to 48% in evaluations of adult ADHD (Musso & Gouvier, 2014). Incentives for receiving an ADHD diagnosis in early adulthood may include academic and occupational accommodations (e.g., see Harrison, Edwards, & Parker, 2007), as well as psychostimulant medication (DeSantis, Noar, & Webb, 2008). Further, a significant body of literature makes clear that it is relatively easy for adults to feign or exaggerate ADHD symptoms and/or complete neuropsychological measures in a manner that would suggest ADHD (e.g., see Conti, 2004; Molina & Sibley, 2014; Pazol & Griggins, 2012; Marshall, Hoelzle, Heyerdahl, & Nelson, 2016).

Though it is becoming standard clinical practice to administer performance and symptom validity tests (PVTs and SVTs, respectively) to detect symptom feigning or amplification (Bush et al., 2005; Heilbronner et al., 2009), much of the adult ADHD research conducted to date making use of archival clinical datasets have failed to systematically evaluate validity issues. The degree to which consideration of response validity would change research findings is unclear; however, it is certainly plausible that the collective understanding of adult ADHD and the psychometric properties of measures may be meaningfully impacted. As an example, while it is commonly believed that

ADHD and a comorbid mood condition result in more significant neuropsychological impairment than either condition independently (e.g. see Larochette, Harrison, Rosenblum, & Bowie, 2011; Roy, Oldehinkel, & Hartman, 2016), this pattern of test findings did not emerge after excluding patients suspected of engaging in symptom amplification (Hoelzle et al., under review).

The second and equally challenging issue in understanding the psychometric properties of adult ADHD clinical instruments is related to the assumption that childhood and adult ADHD are similar clinical conditions. Under this assumption, similarly structured self-report measures are equally applicable to both populations. However, many researchers have posited that the presentation of ADHD may differ across the lifespan, even proposing alternative diagnostic criteria (e.g., see Ward, Wendar, & Reimherr, 1993; Wender, Wolf, & Wasserstein, 2006). Some claim that cognitive symptoms associated with adult ADHD are fundamentally different than those associated with the disorder during childhood (executive dysfunction versus inattention; Barkley, Murphy, & Fischer, 2008), and some symptoms may only capture childhood experiences (i.e. “driven by a motor”).

The psychometric properties of ADHD measures can be examined at different levels, analysis of scales as a whole and analysis of item level properties. Consideration of factor analytic research allows one to better understand whether meaningful differences are present between child and adult ADHD symptom reporting (and hence the psychometric properties of self-report measures). Invariant structures across the lifespan would suggest a similarity whereas discrepant structures could be interpreted as suggesting that adult and child ADHD are distinct (but possibly related) conditions. This

literature base includes conflicting results. For example, Willcutt and colleagues (2012) reviewed numerous confirmatory and exploratory factor analyses in children and adults. A robust two-factor structure of inattention and hyperactivity/impulsivity was reliably identified underlying observer- (parent; teacher) and child-report ADHD rating forms. Factor structures of adult ADHD rating scales were similar. This review suggests a similarity in how symptoms emerge and co-vary in adults and children, and therefore suggests that child and adult self-report ADHD measures are likely to have similar psychometric properties.

In contrast to Willcutt and colleagues' (2012) conclusion that factor structures are largely invariant across the lifespan, it is noteworthy that many adult ADHD researchers have identified the presence of a three-factor adult ADHD structure and propose that hyperactivity and impulsivity are distinct constructs (Barkley, Murphy, & Fischer, 2008; Span, Earleywine, & Strybel, 2002). Three factor structures have also been observed that consist of executive functioning, inattention/hyperactivity, and impulsivity (Kessler et al., 2010). Overall, these findings suggest the possibility of important differences between ADHD in childhood and adulthood, which supports further investigating of the psychometric properties of adult self-report scales.

In addition to research documenting how symptoms co-vary (i.e., investigation of relevant underlying constructs), researchers have focused their attention on understanding specific relationships between test items (e.g., is a symptom present or not) and latent constructs (e.g., inattention) using item response theory (IRT; Embretson, & Reise, 2013; Reise & Waller, 2009). Briefly, IRT allows researchers to (1) evaluate how well an item, reflecting a symptom, represents a latent trait, (2) its ability to discriminate between high

and low levels of a latent trait, and (3) its likelihood of endorsement (i.e., symptoms are not uniformly associated with a latent trait). Thus, IRT analyses allow for a complex analysis of ADHD self- and observer-report measure item functioning. Most of the research conducted to understand how ADHD behavioral checklist items function has made use parent and teacher ADHD rating scales (e.g., Gomez, 2008a; Gomez, 2008b; Li, Reise, Chronis-Tuscano, Mikami, & Lee, 2016; Makransky & Bilenberg, 2014; Purpura, Wilson, & Lonigan, 2010).

While IRT results make clear that ADHD rating scale test items are meaningfully related to theoretical constructs of inattention and hyperactivity/impulsivity, item-level analyses reveal that items of self and observer report measures function discrepantly. As an example, Gomez (2008a) used IRT to evaluate symptom endorsement of ADHD and latent traits of inattention and hyperactivity in elementary-aged children. Overall, parent and teacher ratings of ADHD symptoms were good discriminators of respective latent traits of inattention and hyperactivity/impulsivity. Nevertheless, there were notable differences in how specific items functioned. For example, the inattentive symptom "loses necessary things" was less discriminative than "attention," which means the former symptom is more likely to be endorsed by individuals observing children with higher and lower levels of inattention whereas the latter symptom is likely to be endorsed by individuals observing only children with higher levels of inattention. In contrast to studies investigating elementary school students (Gomez, 2008a; 2008b), Purpura and colleagues (2010) reported that the item "losing necessary things" effectively discriminated between preschoolers with high and low levels of inattention. Findings such as this could suggest that the diagnostic symptom "loses necessary things" is a

common childhood behavior and may not be consistently associated with the latent trait of inattention in elementary-aged children. However, this item may provide more information in preschool-aged children.

This body of IRT literature also suggests some redundancy between select ADHD items and associated relationships with theoretical constructs (i.e., items have comparable threshold parameters). For example, items "difficulty awaiting turn" and "fidgets or squirms" are similarly related to the construct hyperactivity and impulsivity, and therefore may provide redundant information when quantifying this trait (e.g., see Purpura et al., 2010). Additionally, the hyperactivity/impulsivity items "talks excessively" and "blurts out answers" also provide redundant information (Gomez, 2008a; Purpura et al., 2010), and the removal of either item would not reduce measurement precision (Li et al., 2016).

The child and observer IRT literature suggests there is evidence that ADHD rating scale items function in different ways and a similar raw symptom count could reflect vastly different amounts of latent inattention or hyperactivity/impulsivity between individuals. ADHD symptoms, represented by items on behavior rating scales, are not psychometrically equivalent and certain symptoms may deserve greater weight, potentially leading to more accurate diagnosis (Li et al., 2016).

IRT analyses of adult self-report measures are limited, and there have been no attempts to understand item level functioning of retrospective ratings of childhood ADHD symptoms. Gomez (2011) conducted analysis of Barkley's Adult ADHD Rating Scale-Current Symptom Scale (CSS; Barkley & Murphy, 2006b), utilizing a large normative sample. Gomez concluded that all symptoms were relatively good

discriminators of respective latent traits inattention, hyperactivity, and impulsivity. More specifically, inattention symptoms "doesn't listen when spoke to" and "loses things necessary for tasks" were less effective at discriminating between adults with high and low levels of inattention relative to other inattentive symptoms. This finding, which indicates that items differ in their relationship with latent trait of inattention, is not surprising given the frequency of ADHD symptom endorsement across samples. Indeed, survey findings document that at least approximately 25% to 45% of non-clinical samples of adults endorse experiencing ADHD symptoms on self-report measures (DuPaul et al., 2001; Murphy & Barkley, 1996; Gomez, 2011).

Notably, Gomez (2011) evaluated hyperactivity and impulsivity items as separate measures, which contrasts with the *Diagnostic Statistical Manual of Mental Disorders (DSM-5;* APA, 2013) diagnostic structure that specifies hyperactivity/impulsivity as a single construct with an ADHD diagnosis. Gomez reported hyperactivity items "fidgets with hands and feet" and "difficulties with leisure activities" emerged with discriminative parameters similar to inattentive items of "doesn't listen when spoken to" and "loses things necessary for tasks", and were thus less effective as discriminating high and low hyperactivity traits. However, items associated with the latent trait of impulsivity, such as "blurts out answer before question" and "difficulty awaiting turn" were identified as effectively discriminating. Thus, it is unclear how items function within the two-factor structure presented in *DSM-5*.

In summary, understanding the psychometric properties of adult ADHD rating scales is challenging due to an emerging evidence base proposing that ADHD symptoms are not psychometrically equivalent. A significant portion of this research has primarily

investigated parent and teacher observations of ADHD symptoms in children and may not be relevant to understanding ADHD in adults. Furthermore, the adult research literature is limited and has only made use of one normative sample. No research conducted to understand the psychometric studies of ADHD rating scales has considered the validity of symptom reporting or retrospective report of ADHD symptoms in childhood. A greater understanding of the psychometric properties of adult ADHD self-report measures has the potential to improve adult ADHD assessment.

Current Study

There are many challenges associated with assessment and diagnosis of ADHD in adulthood. Failing to consider response validity has the potential to confound interpretation of symptom endorsement and clinical decision-making. Further, very little is known about how self-report measures of ADHD in adulthood represent the theoretical constructs of inattention and hyperactivity/impulsivity. The reported factor structure of adult ADHD self-report measures is inconsistent, and specific test items appear to function in different ways. Thus, there is a need to comprehensively evaluate the psychometric properties of adult ADHD rating scales to improve clinical practice.

The current study evaluated Barkley's Adult ADHD Rating Scale-IV, Self-Report of Current Symptoms (CSS), a self-report diagnostic checklist of current symptoms of ADHD in adults, using a graded response model (GRM) of IRT analysis (Aim 1). A similar self-report measure quantifying retrospective report of childhood symptoms, Barkley's Adult ADHD Rating Scale-IV, Self-Report of Childhood Symptoms (BAARS-C), was also evaluated (Aim 2). Differences in item functioning were also considered after identifying and excluding individuals with suspect effort (Aims 3: CSS-Valid; Aim

4: BAARS-C Valid).

Method

Participants

A retrospective chart review was conducted on 452 adult patients referred to a Midwestern neuropsychology clinic to determine whether they met diagnostic criteria for ADHD. To be included in the present study, each participant must have completed a BAARS current and childhood symptoms self-report measure. It was not necessary to be diagnosed with ADHD. Additionally, individuals who endorsed two response options for a question (N=2) were removed from sample. A total of 400 patients were included out of the 452, comprising the Full group. Some patients skipped questions, occasionally reducing the N for each item. Demographic and descriptive statistics of the are presented in Table 1. The sample consisted primarily of white, young adults with above average intellectual functioning. Consistent with the base rates of ADHD (Willcutt et al., 2012), more men than women comprised this sample. This data has been previously used to investigate frequencies of performance and symptom validity test failure (Marshall et al., 2010; Marshall et al., 2016) and the neuropsychological functioning of individuals with ADHD and/or mood disorders (Hoelzle et al., under review). The Valid Only group is comprised of individuals who were not identified as putting forth suspect effort (*N*= 293).

Individuals identified as putting forth suspect effort during the neuropsychological evaluation were removed for the secondary analyses based on SVT/PVT performance. Insufficient effort was defined as failure on two or more SVT/PVTs (Slick, Sherman, & Iverson, 2010). Performance on the following seven

measures were considered: b Test (e-score of 70 or more, 2 or more commission errors, 2 or more *d* errors, or completion time of 550 or more seconds; Marshall et al., 2010), CVLT-II Forced Choice Recognition (two or more errors; Root, Robbins, Chang, & van Gorp, 2006), Dot Counting Test (e-score of 14 or greater; Marshall et al., 2010), Reliable Digit Span (a score of 6 or less; Babikian, Boone, Lu, & Arnold, 2006), Sentence Repetition (a score of 10 or less; Schroeder & Marshall, 2010), TOVA (total response time variability > 180 ms, 26 or more omission errors, and 31 or more commission errors; Marshall et al., 2010), and Word Memory Test (less than 82.5% correct for immediate recognition, delay recognition, or recall consistency; Green, 2003). Finally, the battery included one SVT, the Clinical Assessment of Attention Deficit-Adult (CAT-A) Infrequency Scale (a score of 3 or greater; Bracken & Boatwright, 2005). Assessment of SVT and PVT performance identified 106 individuals putting forth suspect effort, comprising the Suspect group.

Primary Measures

Barkley's Adult ADHD Rating Scale- Current Symptoms Scale (CSS).

The CSS, which has also been referred to as Barkley's Adult ADHD Rating Scale (BAARS), is an 18-item self-report measure of current ADHD symptoms in adulthood. The CSS was developed directly from DSM-IV symptom criteria with developmentally appropriate verbiage and with each question equating to one specific diagnostic symptom. Nine CSS items represent inattention (IA) symptoms, and the other nine items represent hyperactive/impulsive (H/I) symptoms (6 reflect hyperactivity; 3 reflect impulsivity). CSS items represent potential ADHD symptoms and are rated on a 4 point

Likert scale (0=Not at All, 1=Sometimes, 2=Often, 3=Very Often), and items endorsed as 2 or 3 are considered positive for symptomology. Self-report of ADHD is considered positive if the patient indicates six or more positive endorsements on one or both subscales. Notably, the requirement of six or more positive endorsements is inconsistent with the current DSM-5 diagnostic criteria, which stipulates only five symptoms are required. Additionally, Barkley reported that a Total Score ≥ 1.5 SD's above the sample mean may also be interpreted as reflecting significant ADHD symptomology. Internal consistency of CSS subscales varies from 0.75 to 0.93 (Taylor, Deb, & Unwin, 2011). In the current sample, the CSS IA subscale alpha coefficient was 0.83 and the H/I subscale alpha coefficient was 0.83[1].

Barkley's Adult ADHD Rating Scale- Childhood Symptoms (BAARS-C).

Similar to the CSS, the BAARS-C equates each question to a specific diagnostic criterion. The BAARS-C also contains 18 items, nine of which represent inattentive symptoms in childhood and nine that represent hyperactive/impulsive childhood symptoms (6 reflect hyperactivity; 3 reflect impulsivity). As with the CSS, retrospective report of symptoms in childhood are rated on a 4 point Likert scale (0=Not at All, 1=Sometimes, 2=Often, 3=Very Often), and items endorsed as 2 or 3 are considered positive for childhood ADHD symptomology. The cut-off score is six or more positive endorsements on one or both subscales. A total score ≥ 1.5 SDs above the mean is also considered significant childhood ADHD symptomology. Notably, this is in contrast with the DSM-5, which stipulates that childhood symptoms must be present, but does not

[1] The internal consistency of the CSS with invalid cases removed was α=0.81 for IA and α=0.80 for H/I measures.

specify how many symptoms are necessary. Internal consistency has been reported to range from .88 to .95 (Katz, Petscher, & Welles, 2009; Barkley, 2006). In the current sample, the BAARS-C IA subscale alpha coefficient was 0.88 and the H/I subscale alpha coefficient was 0.87[2].

Data Analytic Plan

Preliminary Analyses.

Mean item scores and frequency of significant item endorsement ("often" or "very often") are reported for the full, valid only, and suspect only samples. Further, to assess potential differences in Valid Only and Suspect sample characteristics, independent sample t-tests were conducted to compare demographic characteristics and symptom endorsement on the CSS and BAARS-C.

Item Response Theory (IRT).

Model selection.

The current sample size is larger than the recommendation of 10 participants per item (336 versus 180; see Brown, 2014), and within the range of sample sizes reported in published literature (n = 105, Mokros et al., 2012; n = 32,000, Reise & Waller, 2003). Of note, following the removal of individuals with insufficient effort (n = 106), sample size decreased (approximately 1/3 of the sample was excluded).

---

[2] The internal consistency of the BAARS-C with invalid cases removed was α=0.87 for IA and α=0.86 for H/I measures.

CSS and BAARS-C item level responses were investigated using IRTPRO (Cai, du Toit, & Thissen, 2011). In clinical contexts, both self-report measures are utilized in a binary, or dichotomous fashion. However, this approach of transforming each item to a dichotomous item (0 or 1 endorsement as no symptomology and 2 or 3 indicative of positive symptomology) is inconsistent with the literature investigating the item functioning of ADHD self- and observer report forms. The IRT model most commonly used is the graded response model (GRM; Samejima, 1969), which accommodates for a polytomous response format (e.g., see Gomez, 2008a; 2011). In brief, GRM develops three response dichotomies for the four CSS and BAARS-C response options: (1) comparing the first category with all others, (2) comparing the first two categories with the last two categories, and (3) comparing the last category with all others. The GRM was selected because it provides more information regarding polytomous item functioning, in addition to providing data relevant to clinical practice (i.e., comparing the first two categories with the last two categories).

In IRT, the probability of endorsing a specific item is related to an underlying latent trait level. All IRT analyses were focused on estimating latent trait levels of inattention and hyperactivity/impulsivity (θ) ranging from 3 SD above to 3 SD below the mean of an assumed normal distribution (M = 0.00, SD = 1.00). Item response function is generally derived from two parameters, item threshold parameters (ß) and item discrimination parameter (α). The former identifies at what trait level there is a 50% probability of endorsing an item. The latter reflects the ability of an item to differentiate individuals at different thresholds (i.e., high versus low inattention). If an item is "easy," individuals with lower and higher levels of a latent trait are likely to endorse the item. In

contrast, if an item is "difficult," only individuals with a higher level of a latent trait are likely to endorse the item.

Unidimensionality.

IRT requires that the scale measure a unidimensional trait. The assumption of unidimensionality is met when a set of data demonstrates a dominant factor that influences item responses (Hambleton, Swaminathan, & Rogers, 1991). While published factor analytic studies suggest two dominant factors underlying these behavioral rating scales (Willcutt et al., 2012), confirmatory factor analyses were conducted to assess unidimensionality of IA and H/I measures. Consistent with a broad literature, inattention and hyperactive/impulsive items were analyzed separately (e.g., see Gomez, 2008a; 2008b; 2011; Purpura et al., 2010). M*plus* (Muthén & Muthén, 2006) was used to conduct confirmatory factor analysis (CFA) to evaluate whether the respective subscales were unidimensional.

A two-factor CFA model, comprised of inattention items (IA) and hyperactivity/impulsivity items (H/I), was assessed using the mean and variance adjusted weighted least squares (WLSMV). First, the two-factor model was fit to both measures of CSS and BAARS-C, using full data. The two-factor model was also fit to both measures following removal of plausibly invalid patient reports. Fit statistics assessed included the chi-square estimates, the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993), the compare fit index (CFI: Bentler, 1990), and the Tucker-Lewis Index (TLI; Bentler, 1990).

Local independence.

IRT analyses also require meeting the assumption of local independence. That is, a response on one item should not impact responses to other items on the measure. Thus, only ability level and item characteristics should influence response. Assessment of the assumption of local independence and IRT analyses were conducted using IRTPRO (Cai, du Toit, & Thissen, 2011). The χ^2 statistics of the observed and expected frequencies in each of the two-way cross tabulations between responses of each item were compared (Chen & Thissen, 1997). Chi-square values are standardized and computed by comparing the observed and expected frequencies in each of the two-way cross tabulations between responses of each item and other items. χ^2 values greater than 10 indicated a violation of the local dependence assumption.

Results

Descriptive Statistics

The Suspect group had significantly lower estimated full-scale IQ (FSIQ) compared to the Valid group (t (395) = -9.38, p <0.001, d =1.04), which is likely due to response distortion on tasks utilized to quantify FSIQ. There were also significant differences in current and retrospective IA and H/I symptom endorsement between the Suspect and Valid groups (See Table 2). Individuals putting forth suspect effort endorsed significantly more IA and H/I symptoms than the valid group (Cohen's d values ≥ .62), and consequently had significantly higher subscale scores (Cohen's d values ≥ .83). Additionally, the mean response and frequency of endorsement of each CSS (Table 3) and BAARS-C (Table 4) item are provided.

Item Response Theory Assumptions

Unidimensionality.

With respect to the full sample, RMSEA values, CFI, and TLI values for the two-factor inattention and hyperactive/impulsive model showed adequate fit for the CSS ($\chi^2(134) = 478.36$, $p < 0.001$, CFI = 0.91, TLI = 0.90, RMSEA = 0.080 (90% CI: [0.07, 0.09]). The CSS factor loadings ranged from 0.54-0.71 for IA and 0.60-0.76 for H/I (See Table 5). The items "easily distracted" and "forgetful in daily activities" had the highest loadings for the IA factor (.71). The item "avoids tasks involving sustained effort" produced the lowest loading (.54). Item "difficulty awaiting turn" had the highest loading for the H/I factor (.76), while "talks excessively" was the lowest loading (.60).

Fit statistics showed adequate fit for the BAARS-C measure ($\chi^2(134) = 602.87$, $p < 0.001$, CFI = 0.92, TLI = 0.91, RMSEA = 0.09 (90% CI: [0.09, 0.10]). The BAARS-C factor loadings ranged from 0.67-0.82 for IA dimension and 0.64-0.81 for H/I dimension (See Table 6). The item "easily distracted" had the highest loading for the IA factor (.82). Items "careless mistakes at work", "difficulty organizing tasks/activities", "avoids tasks involving sustained effort", and "loses things necessary for tasks" comprised the weakest loadings (.67) for the IA factor. Item "difficulty awaiting turn" had the highest loading for the H/I factor (.81), and "feeling on the go" produced the lowest loading (.64).

Based on Chen's (2007) recommendation of comparing models, model fit did not meaningfully change following removal of invalid cases for the CSS ($\chi^2(134) = 354.22$, $p < 0.001$, CFI = 0.91, TLI = 0.90, RMSEA = 0.08 (90% CI: [0.07, 0.09]) or BAARS-C ($\chi^2(134) = 453.45.89$, $p < 0.001$, CFI = 0.93, TLI = 0.92, RMSEA = 0.09 (90% CI: [0.08, 0.10]). Factor loadings ranged from 0.52-0.72 (IA) and 0.54- 0.74 (H/I) for CSS-Valid (see Table 5), and 0.63-0.82 (IA) and 0.60-0.82 (H/I) for BAARS-C-Valid (see Table 6).

Items “avoids tasks involving sustained effort” continued to have the lowest loading (.52) for CSS IA dimension and “easily distracted” and “forgetful in daily activities” continued to have the highest factor loadings (.68 and .72, respectfully) for the CSS H/I dimension. Within the BAARS-C measure, “avoids tasks involving sustained effort” remained the item with the weakest loading (.63) and “difficulty awaiting turn” remained the item with the highest loading (.82).

Local Independence.

The χ^2 statistics of the observed and expected frequencies in each of the two-way cross tabulations between responses of each item were compared (Chen & Thissen, 1997). No standardized χ^2 values were greater than 10.

Graded Response Model

CSS Item discrimination and threshold parameters.

A single discrimination parameter (α), which quantifies the ability of the item to distinguish between higher and lower levels of latent IA or H/I, was obtained for each item. Higher discrimination parameters indicate an item more optimally differentiates between high and low levels of the latent trait. Discrimination estimates for CSS ranged from 1.08 to 2.18 for IA items and 1.19 to 1.95 for H/I items (see Table 7). The most discriminative IA item emerged as “forgetful in daily activities” (α=2.18) and least discriminative item was “avoids tasks” (α=1.08). “Difficulty awaiting turn” was the most discriminative H/I item (α=1.95). The least discriminative H/I item was “fidgets with hands/feet” (α=1.19). The highest and lowest discriminating items did not change following removal of invalid cases (see Table 7).

Threshold parameters (β) for the CSS IA measure are also presented in Table 7. Threshold parameters identify at what trait level there is a 50% probability of endorsing an item at each response category (i.e., endorsement of "(0) Not at All" vs. "(1) Sometimes", "(2) Often", or "(3) Very Often"; 0, 1 vs. 2, 3; or 0, 1, 2 vs. 3). Item "easily distracted" consistently emerged as the lowest threshold for each response dichotomy ($\beta_{1,2,3}$= -4.13, -2.05, -0.51). Item "doesn't listen" consistently emerged with the highest threshold parameters ($\beta_{1,2,3}$= -1.45, 0.60, 2.20). This pattern remained following removal of invalid cases (see Table 8).

Within the H/I measure, "fidgets with hands/feet" consistently emerged as the lowest threshold for each response dichotomy ($\beta_{1,2,3}$= -2.27, -0.99, 0.25), with "feels restless" also having the lowest theta for the first response dichotomy (β_1= -2.27). Item "leaves seat" emerged as highest threshold parameter across all response dichotomies ($\beta_{1,2,3}$= 0.06, 1.38, 2.49). This pattern remained following removal of invalid cases (see Table 10).

BAARS-C Item discrimination and threshold parameters

Discrimination estimates for BAARS-C ranged from 1.55 to 2.50 for IA measure and 1.37 to 2.35 for H/I measure (see Table 9). The most discriminative IA item emerged as "doesn't follow instructions, finish work" (α=2.50), and the lowest discriminating item was "loses things necessary for tasks" (α=1.55). "Difficulty awaiting turn" (α=2.35) emerged as the most discriminating H/I item, and "fidgets with hands/feet" emerged as lowest (α=1.37). While, the item with lowest discrimination changed with removal of invalid cases, from "fidgets with hands and feet" (α=1.37) to "difficulty with leisure activities" (α=1.34), the general pattern was similar across analyses.

Threshold parameters for BAARS-C IA items are also presented in Table 9. Item "easily distracted" consistently emerged as the lowest β for each response dichotomy ($\beta_{1,2,3}$=-2.25, -0.90, -0.26). Item "doesn't listen" consistently emerged as the highest threshold parameters ($\beta_{1,2,3}$= -1.04, 0.55, 1.75), with "doesn't follow instructions" having the highest theta for the first response dichotomy (β_1= -0.86). This pattern remained following removal of invalid cases (see Table 10).

Threshold parameters for H/I items are presented in Table 8. Item "fidgets with hands/feet" consistently emerged as the lowest β for each response dichotomy ($\beta_{1,2,3}$=-2.19, -0.79, 0.45). Item "leaves seat" emerged as highest β_1 parameter (β_1=-0.17). Items "leaves seat" and "difficulty with leisure activities" represented the highest theta values for β_2 and β_3 response categories ("leaves seat", $\beta_{2,3}$=0.80, 1.53; "difficulty with leisure activities", $\beta_{2,3}$=0.79, 1.77). This pattern remained following removal of invalid cases (see Table 10).

Discussion

There are significant challenges associated with assessment ADHD in adulthood. It is increasingly recognized that a significant percentage of adult patients may fabricate or exaggerate ADHD symptoms when completing self-report measures in hopes of securing a diagnosis. Further, there are conflicting findings surrounding the similarity between ADHD presentation in adults and children, reflected in rating-scales and symptoms outlined in the diagnostic criteria. While a significant body of literature documents the psychometric properties of child- and observer-ADHD rating forms, relatively little is known regarding how adult or retrospective childhood ADHD forms function. This research addressed the need to better understand self-report measures

utilized during adult ADHD evaluations. Specifically, a comprehensive item-level analysis of adult ADHD rating scales in a clinical population was conducted providing novel and valuable information for clinicians and researchers.

The aim of this project was to assess the psychometric properties items from of a self-report of ADHD symptoms in adulthood (Barkley's Adult ADHD Rating Scale-Current Symptoms Scale [CSS]) and self-report of symptoms in childhood (Barkley's Adult ADHD Rating Scale-Childhood Symptoms Scale [BAARS-C]) using GRM from IRT. This research builds upon the work of Gomez (2011), who utilized a normative sample to investigate the item level functioning of the CSS. This is the first study to evaluate these scales in a referred clinical sample of adults. Further, this is the first study to conduct CFA and IRT analyses with retrospective self-report of childhood symptoms. Finally, though it is unclear how response and performance validity may impact the psychometric properties of ADHD rating scales, sensitivity analyses were conducted prior to and after carefully considering symptom and performance validity.

Prior to investigating item-level functioning, confirmatory factor analyses were conducted to assess the IRT assumption of unidimensionality. Observed factor structures underlying the CSS and BAARS-C contribute and can be compared to a broad and relevant factor-analytic literature. While many ADHD rating forms reflecting *DSM-5* diagnostic criteria have an underlying two factor structure consisting of inattention and hyperactivity/impulsivity (Willcutt et al., 2012; Taylor, Debb, & Unwin, 2011), this has not always been the case. Additional factor structures have been found (Gomez, 2011). Further, scales that include a wider range of items often have discrepant and more differentiated factor structures (Kessler et al., 2010).

Here, there was strong support for a two-dimensional structure, that discretely emphasized inattention and hyperactivity/impulsivity items on both the CSS and BAARS-C measures. Thus, the current symptom factor structure was similar to a retrospective factor structure, and indirectly provides some support for *DSM-5* specified ADHD presentations. These results are also consistent with many prior investigations that identified separate factors of inattention and hyperactivity/impulsivity of adult ADHD rating scales (e.g. see Willcutt et al., 2012) and supports the decision to analyze inattention symptoms and hyperactivity/symptoms separately. Importantly, Gomez (2011) also conducted CFA prior to conducting IRT and identified three factors representing inattention, hyperactivity, and impulsivity. Discrepant factor structures preclude a direct comparison of findings.

It is noteworthy that more recent factor analytic research supports that a bifactor dimensional structure underlying ADHD self-report measures (reflecting ADHD, hyperactivity/impulsivity, and inattention) may more adequately describe covariance between items than the two-dimensional structure (Li et al., 2016; Matte et al., 2015). Current findings provide tentative support for this approach given that factors of inattention and hyperactivity/impulsivity were significantly correlated. Consistent with Li and colleagues' (2016) methodology, this suggests that multidimensional IRT analysis would have been an appropriate analytic strategy. Not accounting for an association between IA and H/I constructs is a potential limitation of this study; however, given adequate model fit of the replicated two-dimensional structure, and the fact that ADHD is conceptualized clinically as consisting of two independent but related constructs,

analyses assessed IA and H/I items as separate measures to match how the CSS and BAARS-C symptom scales are utilized.

In addition to documenting the factor structure of the CSS and BAARS-C in an adult clinical sample, this research substantively adds to what is known about the item-level functioning of the respective adult ADHD rating forms. IRT analyses have primarily focused on documenting the psychometric properties of children and observer report forms. Across studies, items which appear to reflect either inattention or hyperactivity/impulsivity are not equally related to corresponding constructs. For example, Li and colleagues (2016) reported the symptom "often talks excessively" to be the least and symptoms "attention" to be the most informative. Gomez (2008a) also reported "attention" to be most informative, but "loses" was the least informative. Consideration of findings across studies offers clinicians a more nuanced understanding of how items function and illuminates which items might have the greatest diagnostic utility.

CSS

Consistent with prior IRT analyses of ADHD symptom report forms, the discrimination of specific CSS items varied. Comparison of item discrimination parameters allow clinicians and researchers to better understand which items are likely to differentiate between individuals with high and low latent traits. Within the IA measure, discrimination parameters ranged from 1.08 to 2.18, which is comparative to the range of IA discrimination parameters derived making use of a normative adult population (Gomez, 2011; α =1.32 to 2.12). Specifically, the item "forgetful in daily activities" optimally discriminated between higher and lower levels of latent trait of IA. However, in

contrast, the item "avoids tasks involving sustained effort" was the least discriminative IA item. Despite each of these items reflecting a specific *DSM-5* criterion of ADHD, IRT results reveal that the items "avoids tasks" and "forgetful in daily activities" function very differently in their ability to distinguish those with higher or lower levels of inattentiveness. Avoiding tasks is a commonly reported adult behavior, thus this item is likely capturing a rather non-specific behavior rather than perhaps a more pathological and impairing indication of inattention.

In addition to discrimination parameters, the GRM provides three item thresholds representing the measure's three possible response dichotomies (i.e., endorsement of "(0) Not at All" vs. "(1) Sometimes", "(2) Often", or "(3) Very often"; 0, 1 vs. 2, 3; or 0, 1, 2 vs. 3). Consideration of item threshold parameters allow clinicians and researchers to better understand the 50% likelihood of item endorsement at each response category given an amount of latent trait. For clinical interpretation of this measure, the β_2 item threshold parameters are of particular interest, in that they represent the amount of latent IA needed to endorse the item at a "clinically significant" level (i.e. "often" or "very often"). Interpretation of the CSS IA measure reveals that very little latent IA trait is necessary to have a 50% likelihood of endorsing "easily distracted". In the β_2 response category, individuals with 2 standard deviations below the mean of latent trait IA would have a 50% likelihood of endorsing this symptom as "often" or "very often". Thus, "easily distracted" is likely be frequently endorsed in individuals with subclinical levels of IA. Further, 8 of the 9 IA items emerged with β_2 threshold parameters below the mean. As such, IA symptoms are more likely to be endorsed at a clinically significant level

when an individual has average or lower inattention. Therefore, lower levels of IA are needed to reach diagnostic criteria.

Inattention items which emerged with lower discrimination parameters and extreme threshold values exemplify the nuances associated with item endorsement on ADHD rating scales. Diagnostically, “easily distracted” is given the same weight toward meeting the symptom threshold as item “doesn’t listen when spoken to” which required the most IA θ trait level to have a 50% of endorsement. This item was also identified as “easiest” in an IRT analysis of parent rating scales of ADHD in childhood (Li et al., 2016) and adult report of current symptoms (Gomez, 2011). In this referred clinical sample, frequency analyses report that 90% of patients reported “often” or “very often” “feeling distracted,” so at face value, it may appear that this item is a strong and specific indicator of ADHD psychopathology. However, results of item level analyses indicate that “feeling distracted” is likely to be endorsed across clinical and normative populations. Consequently, in a survey of adults renewing their driver’s license, 19.1% of adults endorsed this item at a clinically significant level (Murphy & Barkley, 1996). Overall, there is converging evidence that this symptom does not function similarly to other ADHD IA symptoms.

Though the range of IA item discrimination parameters are comparable to those observed in an adult normative sample (Gomez, 2011), there are some notable differences. For example, the item “easily distracted” differentiated between high and low IA in a normative sample in a more effective way than in this clinical sample. This may be plausibly explained by differences in base rates of symptom reporting between the two samples. While Gomez recruited participants from a broader community, self-

report scales in this study were completed as a part of a clinical assessment. Nevertheless, this is still a surprising finding, given that item parameters estimated in IRT analyses are posited to be sample independent (Embertson & Reise, 2001). On the other hand, some have observed that item functioning may differ related to variables of sex, race-ethnicity, and age (e.g., see Li, et al., 2016).

With respect to H/I CSS measure, the range of discrimination parameters was similar to those observed among IA items (α= 1.19-1.95). Notably, across all H/I and IA CSS measures, items assessing impulsivity ("blurts out answer", "difficulty awaiting turn", "interrupts/intrudes") produced the highest discrimination parameters in this study. Specifically, the CSS H/I item "difficulty awaiting turn" optimally discriminated between higher and lower levels of hyperactivity and impulsivity in adult patients (α = 1.95). In contrast, the item "fidgets with hands/feet, squirms" poorly discriminated between higher and lower levels of H/I (α = 1.19).

In comparison to IA items, 7 of the 9 H/I item's β_2 threshold parameters were above the mean, which indicates a lower likelihood of endorsement by individuals with lower latent H/I. Thus, a higher level of H/I is needed to endorse a clinically significant level of symptoms. This is not surprising and fits with a broad literature indicating that IA ADHD presentations are more prevalent than H/I presentations in adulthood (Kessler et al., 2010). The item "fidgets with hands/feet, squirms" seems especially problematic. Threshold parameters, as well as discrimination parameters, suggest that this item provides little information regarding latent H/I. It is likely to be endorsed with individuals with lower levels of latent H/I and, relative to other H/I items, poorly distinguishes between adults with higher and lower levels of H/I. Frequency analyses reveal that it is

often significantly endorsed in this clinical sample (71.8%). Further, in a sample of adult drivers, 20.3% of adults significantly endorsed "fidgets with hands/feet" (Murphy & Barkley, 1996). Thus, this item reflects a DSM symptom criterion, but item level analyses suggest it is common for patients and community members to endorse it regardless of trait level.

Findings regarding the H/I items cannot be directly compared to Gomez's work given he analyzed hyperactivity and impulsivity items separately (i.e. hyperactive and impulsivity were assessed as distinct latent traits). However, comparison of current H/I item functioning and childhood item functioning reveals novel information. For example, parent report of preschool behaviors identified two H/I items, "difficulty awaiting turn" and "fidgets with hands/feet, squirms" as providing redundant information (Purpura et al., 2010), whereas in adults, these items function differently. This highlights that the probability of endorsing a specific ADHD symptom changes across the lifespan and suggests important differences in the psychometric properties of child and adult ADHD forms.

BAARS-C

To further explore potential differences in self-reported ADHD symptoms across the lifespan, item level functioning of retrospective report of childhood symptoms was also explored. In addition to assessing the presence of five or more current ADHD symptoms in adulthood, *DSM-5* stipulates that symptoms be present in childhood prior to the age of 12. However, there is no symptom threshold to be met, but rather the general onset of symptoms before the age of 12. Thus, clinicians using self-reports of

retrospective childhood symptoms should be cautiously aware of the frequency and likelihood of symptom endorsement on these measures.

This is the first study to evaluate retrospective report of childhood symptoms, contributing the clinical utility of measures assessing symptom onset prior to age 12 and the conceptualization of ADHD across the lifespan. Surprisingly, though items assess the same symptoms outlined in the DSM, BAARS-C and CSS item discrimination parameters differed which suggests that the same current and retrospective symptom appears to have different relationships with corresponding latent traits. BAARS-C items on both IA and H/I scales tended to be more effective at discriminating between trait presence than current symptom reports (CSS; α range= 1.08-2.18, BAARS-C α= 1.37 - 2.50). Thus, the ability of items to distinguish between higher and lower levels of IA and H/I differs whether symptoms are retrospectively reported or currently experienced. However, H/I items associated with impulsivity (e.g. “blurts out answers”, “difficulty awaiting turn”, “interrupts”) continued to be most effective in differentiating among patients at varying levels of the H/I trait in both CSS and BAARS-C. These items appear to be most effective in both the CSS and BAARS-C measures, perhaps suggesting further and more critical examination by clinicians. Comparison of threshold parameters between the CSS and BAARS-C were similar, though more (4/9) IA items β_2 threshold parameters fell above the mean. This suggests more latent IA trait is needed to report retrospective IA symptoms compared to current symptoms.

The variability in item functioning between CSS and BAAR-C IA and H/I measures contributes to ongoing discussion of differences between child and adult symptoms of ADHD. Differences in item functioning may be explained by not accurately

recalling childhood experiences (Mannuzza et al., 2002) or a change in how ADHD presents throughout the lifespan. The broader IRT literature of ADHD rating scales, conducted with parent and teacher report of preschool and school-aged children confirm that items function differently across the lifespan. Future work may consider how latent traits of inattention and hyperactivity/impulsivity change within differing developmental contexts. Item level analysis within longitudinal study of children with ADHD followed into adulthood would help solidify the understanding of latent trait stability through development. Additionally, differences between current and retrospective report or observation of childhood behavior suggests that latent traits change during development and should be further studied. A better understanding of these changes might inform substantive changes to adult ADHD diagnostic criteria.

Symptom Validity

Symptom exaggeration and response distortion are important issues to consider when conducting adult ADHD evaluations given increased awareness of ADHD symptomology and incentives for receiving a diagnosis in adulthood. While response validity is increasingly evaluated in research and clinical contexts, it has not been considered in item-level analyses of ADHD self-report measures. In addition to offering a comprehensive understanding of CSS and BAARS-C item functioning, this research investigated the possibility that individuals' attempts to feign or exaggerate ADHD symptoms might alter psychometric findings. Analyses were repeated after removal of 106 patients suspected of putting forth insufficient effort during their neuropsychological evaluation.

It was anticipated that findings would change following the removal of invalid cases. Plausibly, given a general over-reporting of symptoms in the full sample, it was expected that items would function more similarly when all participants were investigated. However, in contrast to hypothetical expectation, our results suggest that items function similarly following removal of data obtained from patients putting forth insufficient effort. The majority of discrimination parameters slightly decreased from the full to valid only analyses on the CSS IA and H/I measures. Thus, screening for insufficient effort did not meaningfully change items abilities to distinguish higher and lower levels of IA and H/I. Threshold parameters slightly increased in the valid-only analyses, which logically follows the need for more latent IA and H/I to meet thresholds of endorsement. The BAARS-C analyses showed more fluctuation from full to valid-only analyses, particularly within the discrimination parameters. This may be related to recall of less specific ADHD symptoms, but rather a syndrome of ADHD in childhood. Overall, discrimination parameters decreased within BAARS-C IA and H/I items. Like the CSS, threshold β estimates increased following removal of individuals with suspect performance.

The similarity between item functioning in both valid only and full samples is in contrast with the findings derived from other studies which strongly support the importance of assessing for valid performance (Edmundson et al., 2017; Smith, Cox, Mowle, & Edens, 2017). Notably, this analysis comprehensively screened for insufficient effort by requiring failure of multiple SVT/PVTs, and thus primarily captured performance distortion rather than symptom exaggeration. Importantly, applying a different insufficient effort criterion may change results and interpretations of item-level

functioning. Further, the self-report measures investigated do not include embedded SVTs to detect exaggerated report of ADHD symptoms. To further investigate whether psychometric properties differ after identifying insufficient effort, the Conner's Adult ADHD Rating Scale (CAARS; Conners, Erhardt, & Sparrow, 1998) should be investigated. Uniquely, the CAARS includes two embedded measures to detect relevant non-credible report of ADHD symptoms, the Infrequency Index (CII; Suhr, Buelow, & Riddle, 2011) and the Exaggeration Index (Harrison & Armstrong, 2016). Despite similar findings before and after excluding participants suspected of insufficient effort, it is nevertheless important for future research related to symptom validity in ADHD assessment.

Theoretical and Clinical Implications

A comprehensive evaluation of how specific ADHD symptoms relate to the theoretical constructs of inattention and hyperactivity/impulsivity in adult clinical samples is warranted, given disagreements surrounding its presentation in adulthood (Riccio et al., 2005; Faraone, Biederman, Mick, E. 2006; Faraone & Biederman, 2016). As items closely reflect diagnostic criteria, a greater understanding of items and symptoms permits for a more tangible, quantitative grasp of ADHD psychopathology in adulthood. The application IRT to direct item-to-symptom measures allows for a unique psychometric assessment of how the current *DSM-5* symptoms represent latent traits of inattention and hyperactivity/impulsivity.

Overall, these data suggest that CSS and BAARS-C items generally reflect latent traits of ADHD, though in different ways. Notably, the item "easily distracted" appears to perform poorly across current report of symptoms and retrospective childhood symptoms.

While “easily distracted” is a hallmark feature of ADHD, it is problematic that many people report this experience. Clinicians may further inquire about functional and domain specific impairment when this symptom is endorsed to ensure a true clinically significant level of distress is present. In contrast, items “blurts out answers”, “difficult awaiting turn” and “interrupts/intrudes” appear to uniquely capture H/I ADHD presentations.

Given the use of a symptom count in ADHD as a categorical approach to diagnosis, the importance of symptoms accurately and uniquely capturing ADHD traits cannot be understated, particularly as the symptom threshold has been lowered from six to five for adults in the *DSM-5*. Diagnostically, symptoms carry equal weight, but these results suggest that they differ in likelihood of endorsement and their ability to differentiate across the latent trait continuum. It is debatable whether all symptoms should be given equal weight when formulating symptom counts, as they differ in likelihood of endorsement by individuals with subclinical ADHD.

Additionally, with the ADHD diagnostic criteria requiring symptom onset prior to age 12, careful consideration should be given to how the likelihood of symptom endorsement of IA and H/I changes throughout development. Indeed, results from these data suggest that adult IA items function differently than retrospective childhood IA items, particularly in their ability to discriminate higher and lower levels of latent inattention.

The use of self-report measures with items that directly parallel diagnostic criteria for ADHD comes with some trade-offs. While these measures directly assess significant symptom presence and unambiguously quantify symptom thresholds, these data indicate that significant endorsement of one item is not equivalent to significant endorsement of

another. Many adults are likely to acknowledge being “easily distracted”, wherein only four more IA symptoms are needed to reach the diagnostic threshold. Future research is needed to evaluate if clinical practice is improved by utilizing additional ADHD measures. For example, alternative measures ask patients to quantify a broader range of behaviors associated with ADHD (e.g., CAARS; Barkley Deficits in Executive Functioning Scale [BDEFS], Barkley, 2011a) or to indicate how ADHD symptoms impact activities of daily living (e.g., Barkley Functional Impairment scale, Barkley 2011c). Utilizing these scales during clinical assessment may alleviate the limitations of the sole use of symptom checklists.

Future Directions

Clinically, it may be beneficial to utilize an ADHD self-report measure with numerous items related to a symptom in the diagnostic criteria, particularly symptoms that are less likely to be endorsed by individuals with lower trait levels. For example, these data suggest that items capturing impulsivity are the best discriminators of H/I, however adults are less likely to be diagnosed with the H/I ADHD subtype (Kessler et al., 2010). Currently, there are a limited number of symptoms assessing impulsivity, and self-report rating scales for adults may benefit from more items capturing impulsivity. Indeed, factor analytic studies which include more items related to executive functioning (i.e. inhibition, impulsivity) revealed three factor structures of ADHD measures, emphasizing the inclusion of additional symptoms capturing executive functioning be considered for future diagnostic criteria (Kessler et al., 2010).

Additionally, IRT analyses may be utilized to develop adaptive adult ADHD testing paradigms, wherein endorsement of one item leads to the presentation of

additional items related to the same construct at respective trait levels. Indeed, work utilizing this methodology is in the nascent stages of development (e.g., see Ustun et al., 2017). Finally, as presented above, further consideration of differential item functioning analyses in diverse samples may reveal differences related to sample characteristics (Li & Reise, 2016). Some items may have greater utility in different age, racial/ethnic, gender, or urban vs. rural populations due to cultural appraisal of behaviors. Indeed, this sample comprised of mostly white, well-educated, and intelligent patients, so it is unclear how items may function differently given diverse sample characteristics.

Conclusion

Diagnosis of ADHD in adulthood presents clinicians with complex challenges. Response validity may confound interpretation of assessment data, and it is increasingly evident that many individuals engage in response distortion. Additionally, the historical view of ADHD as a childhood condition offers a convoluted path for understanding its presentation in adulthood. This research provides novel information regarding relationships between common adult ADHD self-report form items and corresponding theoretical constructs, which has the potential improve clinical practice. Symptoms of inattention and hyperactivity/impulsivity are endorsed differently across the lifespan, and these data suggest that they vary in their relationship to the theoretical constructs of IA and H/I. At face value, meeting a symptom threshold of five or more symptoms may be misleading. Closer attention given to specific symptoms in the context of the clinical interview and reported difficulties across domains may lead to more informed diagnosis. Though screening for sufficient effort did not meaningfully change item level functioning, it is still important to consider in all adult ADHD evaluations.

References

American Academy of Pediatrics (2011). ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/ Hyperactivity Disorder in Children and Adolescents. *Pediatrics, 128*(5). doi:10.1542/peds.2011-2654

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

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist, 20*(1), 145-159.

Barkley, R.A. (2006a). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment. (*3rd ed.). New York: Guilford

Barkley, R. A., & Murphy, K. (2006b). *Attention deficit hyperactivity disorder: A clinical workbook* (3rd ed.). New York: Guilford Press.

Barkley, R., Murphy, K., & Fischer, M. (2008). *ADHD in Adults: What the Science Says.* New York, New York: The Guilford Press.

Barkley, R. A. (2011a). Barkley deficits in execute functioning (BDEFS). New York: The Guilford Press.

Barkley, R. A. (2011b). Barkley Adult ADHD Rating Scale-IV (BAARS-IV). New York: The Guilford Press.

Barkley, R. A. (2011c). Barkley functional impairment scale (BFIS). New York: Guilford Press.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin, 107*(2), 238.

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

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. & Long, J.S. [Eds.] *Testing structural equation models*. Newbury Park, CA: Sage, 136–162.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., ... & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*(4), 419-426.

Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO for Windows [Computer software]. Lincolnwood, IL: Scientific Software International.

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265-289.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, *14*(3), 464-504.

Conti, R. P. (2004). Malingered ADHD in adolescents diagnosed with conduct disorder: A brief note. *Psychological reports*, *94*(3), 987-988

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

DeSantis, A., Noar, S. M., & Webb, E. M. (2009). Nonmedical ADHD stimulant use in fraternities. *Journal of Studies on Alcohol and Drugs*, *70*(6), 952-954.

DuPaul, G. J., Schaughency, E. A., Weyandt, L. L., Tripp, G., Kiesner, J., Ota, K., & Stanish, H. (2001). Self-report of ADHD symptoms in university students: Cross-gender and cross-national prevalence. *Journal of learning disabilities*, *34*(4), 370-379.

Edmundson, M., Berry, D. T., Combs, H. L., Brothers, S. L., Harp, J. P., Williams, A., ... & Scott, A. B. (2017). The effects of symptom information coaching on the feigning of adult ADHD. *Psychological assessment*, *29*(12), 1429.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Faraone, S. V, & Biederman, J. (2005). What Is the Prevalence of Adult ADHD? Results of a Population Screen of 966 Adults. *Journal of Attention Disorders*, *9*(2), 384–391. https://doi.org/10.1177/1087054705281478

Faraone, S.V., Biederman, J., Mick, E. (2006) The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychological Medicine, 36*(2), 159-165.

Faraone, S. V., & Biederman, J. (2016). Can attention-deficit/hyperactivity disorder onset occur in adulthood?. *JAMA psychiatry*, *73*(7), 655-656.

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

Gomez, R. (2008a). Item response theory analyses of the parent and teacher ratings of the DSM-IV ADHD rating scale. *Journal of Abnormal Child Psychology*, *36*(6), 865-885.

Gomez, R. (2008b). Parent ratings of the ADHD items of the disruptive behavior rating scale: Analyses of their IRT properties based on the generalized partial credit model. *Personality and Individual Differences*, *45*(2), 181-186.

Gomez, R. (2011). Item response theory analyses of adult self-ratings of the ADHD symptoms in the Current Symptoms Scale. *Assessment*, *18*(4), 476-486.

Haavik, J., Halmøy, A., Lundervold, A. J., & Fasmer, O. B. (2010). Clinical assessment and diagnosis of adults with attention-deficit/hyperactivity disorder. *Expert review of neurotherapeutics*, *10*(10), 1569-1580.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

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.

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

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the Neuropsychological Assessment of Effort, Response Bias, and Malingering. *The Clinical Neuropsychologist*, *23*(7), 1093–1129. https://doi.org/10.1080/13854040903155063

Hoelzle, J. B., Ritchie, K., Marshal, P., Vogt, E., & Marra, D. (under review). Erroneous conclusions: The impact of failing to identify invalid symptom presentation when conducting adult attention-deficit/hyperactivity disorder (ADHD) research. *Psychological Assessment.*

Katz, N., Petscher, Y., & Welles, T. (2009). Diagnosing attention-deficit hyperactivity disorder in college students: An investigation of the impact of informant ratings on diagnosis and subjective impairment. *Journal of Attention Disorders*, *13*(3), 277-283.

Kessler, R. C., Adler, L., Barkley, R., Biederman, J., Conners, C. K., Demler, O., … Zaslavsky, A. M. (2006). The prevalence and correlates of adult ADHD in the United States: Results from the National Comorbidity Survey Replication. *The American Journal of Psychiatry*, *163*(4), 716–723. http://doi.org/10.1176/appi.ajp.163.4.716

Kessler, R. C., Green, J. G., Adler, L. A., Barkley, R. A., Chatterji, S., Faraone, S. V., … Brunt., D. L. Van. (2010). Structure and diagnosis of adult attention-deficit/hyperactivity disorder: Analysis of expanded symptom criteria from the adult adhd clinical diagnostic scale. *Archives of General Psychiatry*, *67*(11), 1168–1178. Retrieved from http://dx.doi.org/10.1001/archgenpsychiatry.2010.146

Kooij, J. J. S., Buitelaar, J. K., van den Oord, E. J., Furer, J. W., Rijnders, C. A. T., & Hodiamont, P. P. G. (2005). Internal and external validity of attention-deficit hyperactivity disorder in a population-based sample of adults. *Psychological Medicine*, *35*(6), 817–27. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15997602

Kooij, S. J. J., Boonstra, M. A., Swinkels, S. H. N., Bekker, E. M., de Noord, I., & Buitelaar, J. K. (2008). Reliability, Validity, and Utility of Instruments for Self-Report and Informant Report Concerning Symptoms of ADHD in Adult Patients. *Journal of Attention Disorders*, *11*(4), 445–458. https://doi.org/10.1177/1087054707299367

Larochette, A. C., Harrison, A. G., Rosenblum, Y., & Bowie, C. R. (2011). Additive neurocognitive deficits in adults with attention-deficit/hyperactivity disorder and depressive symptoms. *Archives of clinical neuropsychology*, acr033.

Li, J. J., Reise, S. P., Chronis-Tuscano, A., Mikami, A. Y., & Lee, S. S. (2016). Item Response Theory Analysis of ADHD Symptoms in Children with and without ADHD. *Assessment*, *23*(6), 655–671. https://doi.org/10.1177/1073191115591595

Makransky, G., & Bilenberg, N. (2014). Psychometric Properties of the Parent and Teacher ADHD Rating Scale (ADHD-RS): Measurement Invariance Across Gender, Age, and Informant. *Assessment*, *21*(6), 694–705. https://doi.org/10.1177/1073191114535242

Mannuzza, S., Klein, R. G., Klein, D. F., Bessler, A., & Shrout, P. (2002). Accuracy of adult recall of childhood attention deficit hyperactivity disorder. *American Journal of Psychiatry*, *159*(11), 1882-1888.

Matte, B., Anselmi, L., Salum, G. A., Kieling, C., Gonçalves, H., Menezes, A., … Rohde, L. A. (2015). ADHD in DSM-5: a field trial in a large, representative sample of 18- to 19-year-old adults. *Psychological Medicine*, *45*(2), 361–373. https://doi.org/DOI: 10.1017/S0033291714001470

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

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

Moffitt, T. E., Houts, R., Asherson, P., Belsky, D. W., Corcoran, D. L., Hammerle, M., … Caspi, A. (2015). Is adult ADHD a childhood-onset neurodevelopmental disorder? Evidence from a four-decade longitudinal cohort study. *American Journal of Psychiatry*, *172*(10), 967–977. https://doi.org/10.1176/appi.ajp.2015.14101266

Mokros, A., Schilling, F., Eher, R., & Nitschke, J. (2012). The Severe Sexual Sadism Scale: cross-validation and scale properties. *Psychological Assessment*, *24*(3), 764.

Molina, B. S., & Sibley, M. H. (2014). The case for including informant reports in the assessment of adulthood ADHD. *The ADHD Report*, *22*(8), 1-7.

Murphy, K., & Barkley, R. A. (1996). Prevalence of DSM-IV symptoms of ADHD in adult licensed drivers: Implications for clinical diagnosis. *Journal of Attention Disorders*, *1*(3), 147–161. https://doi.org/10.1177/108705479600100303

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

Muthén, L. K., & Muthén, B.O. *Mplus User's Guide*. 6th edition. Muthén & Muthén; Los Angeles, CA: 1998-2011

Purpura, D. J., Wilson, S. B., & Lonigan, C. J. (2010). Attention-deficit/hyperactivity disorder symptoms in preschool children: Examining psychometric properties using item response theory. *Psychological Assessment*, *22*(3), 546-558.http://dx.doi.org/10.1037/a0019581

Pazol, R., & Griggins, C. (2012). Making the case for a comprehensive ADHD assessment model on a college campus. *Journal of College Student Psychotherapy*, *26(1*), 5-21.

Post, R. E., & Kurlansik, S. L. (2012). Diagnosis and management of attention-deficit/hyperactivity disorder in adults. *American family physician*, *85*(9), 890.

Reise, S. P. (1990). A comparison of item-and person-fit methods of assessing model-data fit in IRT. *Applied Psychological Measurement*, *14*(2), 127-137.

Reise, S. P., & Waller, N. G. (2003). How many IRT parameters does it take to model psychopathology items?. *Psychological Methods*, *8*(2), 164.

Reise, S. P., & Waller, N. G. (2009). Item Response Theory and Clinical Measurement. *Annual Review of Clinical Psychology*, *5*, 27-48.f

Riccio, C. A., Wolfe, M., Davis, B., Romine, C., George, C., & Donghyung, L. (2005).

Attention deficit hyperactivity disorder: Manifestation in adulthood. *Archives of Clinical Neuropsychology, 20,* 249-269.

Root, J. C., Robbins, R. N., Chang, L., & Van Gorp, W. G. (2006). Detection of inadequate effort on the California Verbal Learning Test-: Forced choice recognition and critical item analysis. *Journal of the International Neuropsychological Society*, *12*(5), 688-696

Ross, D.M., & Ross, S.A. (1976). Hyperactivity: Research, theory, and action. New York: John Wiley.

Roy, A., Oldehinkel, A. J., & Hartman, C. A. (2016). Cognitive functioning in adolescents with self-reported ADHD and depression: results from a population-based study. *Journal of abnormal child psychology*, 1-13.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 43(100).*

Schroeder, R. W., & Marshall, P. S. (2010). Validation of the Sentence Repetition Test as a measure of suspect effort. *The Clinical Neuropsychologist*, *24*(2), 326-343.

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.

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

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 LP-211. Retrieved from http://bjp.rcpsych.org/content/194/3/204.full

Span, S. A., Earleywine, M., & Strybel, T. Z. (2002). Confirming the Factor Structure of Attention Deficit Hyperactivity Disorder Symptoms in Adult, Nonclinical Samples. *Journal of Psychopathology and Behavioral Assessment*, *24*(2), 129–136. https://doi.org/10.1023/A:1015396926356

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

Suhr, J. A., & Berry, D. T. (2017). The importance of assessing for validity of symptom report and performance in attention deficit/hyperactivity disorder (ADHD): Introduction to the special section on noncredible presentation in ADHD. *Psychological assessment*, *29*(12), 1427.

Taylor, A., Deb, S., & Unwin, G. (2011). Scales for the identification of adults with attention deficit hyperactivity disorder (ADHD): a systematic review. *Research in Developmental Disabilities, 32*(3), 924-938.

Ustun, B., Adler, L. A., Rudin, C., Faraone, S. V., Spencer, T. J., Berglund, P., ... & Kessler, R. C. (2017). The World Health Organization adult attention-deficit/hyperactivity disorder self-report screening scale for DSM-5. *Jama psychiatry*, *74*(5), 520-526.

Ward, M. F., Wendar, P. H., & Reimherr, F. W. (1993). The Wender Utah Rating Scale: an aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *American Journal of Psychiatry*, *150*(6), 885–890. https://doi.org/10.1176/ajp.150.6.885

Wender, P. H., Wolf, L. E., & Wasserstein, J. (2006). Adults with ADHD: An overview. *Annals of the New York Academy of Sciences*, *931*(1), 1–16. https://doi.org/10.1111/j.1749-6632.2001.tb05770.x

Willcutt, E. G., Nigg, J. T., Pennington, B. F., Solanto, M. V., Rohde, L. A., Tannock, R., … Lahey, B. B. (2012). Validity of DSM-IV attention–deficit/hyperactivity disorder symptom dimensions and subtypes. *Journal of Abnormal Psychology*, *121*(4), 991–1010. http://doi.org/10.1037/a0027347

Table 1. Demographic information for full, valid-only, and suspect samples.

	Full	Valid Only	Suspect
N	400	293	106
% Male	61.50	63.10	57.50
% Caucasian	76.00	77.80	71.70
Age	26.34 (7.62)	26.23 (7.44)	26.67 (8.13)
Education Years	14.45 (1.69)	14.63 (1.55)	13.93 (1.95)
Estimated Full Scale IQ (FSIQ)	112.51 (16.37)	116.83 (14.39)*	100.95 (16.02)*

*Denotes significant differences at $p < 0.05$.

Table 2. Independent Samples t-Test between Suspect and Valid-Only group.

	Full		Valid Only		Suspect			
Items endorsed**	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*t*	d
CSS-IA	392	5.50 (2.43)	288	5.15 (2.45)	103	6.54 (2.02)	5.68*	0.62
CSS-H/I	396	3.71 (2.46)	289	3.17 (2.22)	106	5.20 (2.50)	7.36*	0.86
BAARS-C-IA	386	4.69 (2.77)	282	4.22 (2.74)	103	6.02 (2.38)	6.30*	0.69
BAARS-C- H/I	390	4.11(2.82)	285	3.60 (2.75)	104	5.52 (2.51)	6.22*	0.73
Total Sum Scores								
CSS-IA	400	27.24 (9.17)	293	25.22 (8.26)	106	33.08 (8.82)	8.24*	0.91
CSS- H/I	400	26.67 (9.26)	293	24.63 (8.30)	106	32.55 (9.05)	8.22*	0.90
BAARS-C IA	398	26.06 (10.80)	291	23.90 (10.21)	106	32.21 (9.81)	7.25*	0.83
BAARS-C H/I	398	25.72 (10.95)	291	23.50 (10.29)	106	32.03 (10.06)	7.35*	0.84

*Denotes significant differences at p <0.05. **Number of items endorsed at a clinically significant level (2= "Often" or 3= "Very Often").

Note. t-test and Cohens d represent comparison of Valid Only and Suspect group means.

Table 3. Descriptive information of item level data of CSS (Mean, Standard Deviation, % significantly endorsed).

	Full			Valid Only			Suspect Only		
	N	M(SD)	% endorsed	N	M(SD)	% endorsed	N	M(SD)	% endorsed
Inattention Symptoms									
1. Careless mistakes at work	400	1.71(0.87)	56.50	293	1.58(0.85)	50.20	106	2.09(0.82)	74.50
2. Poor sustaining attention for task	397	1.78(0.89)	61.00	291	1.71(0.88)	58.40	105	2.01(0.9)	68.90
3. Doesn't listen when spoken to	399	1.25(0.86)	35.30	293	1.13(0.82)	28.70	105	1.64(0.88)	53.80
4. Doesn't follow instructions, finish work	398	1.58(1.02)	50.50	292	1.47(1.01)	45.70	105	1.93(0.97)	64.20
5. Difficulty organizing tasks/activities	400	1.90(0.95)	66.00	293	1.84(0.94)	64.20	106	2.09(0.94)	71.70
6. Avoids tasks involving sustained effort	400	2.02(0.93)	70.30	293	1.93(0.93)	67.20	106	2.30(0.84)	79.20
7. Loses things necessary for tasks	400	1.67(1.02)	53.80	293	1.57(1.01)	49.50	106	1.95(0.97)	66.00
8. Easily distracted	399	2.52(0.69)	90.00	292	2.44(0.71)	88.10	106	2.78(0.50)	96.20
9. Forgetful in daily activities	399	1.92(0.90)	65.30	292	1.86(0.88)	62.80	106	2.13(0.91)	72.60
Hyperactivity/Impulsivity Symptoms									
10. Fidgets with hands/feet, squirms	399	2.06(1.01)	71.80	292	1.97(1.03)	68.60	106	2.34(0.87)	81.10
11. Leaves seat when seating is expected	400	0.75(0.92)	19.50	293	0.59(0.79)	13.70	106	1.19(1.11)	35.80
12. Feels restless	400	1.90(0.92)	66.00	293	1.79(0.91)	60.10	106	2.24(0.87)	83.00
13. Difficulties with leisure activities	400	1.12(0.99)	29.80	293	0.95(0.09)	23.90	106	1.58(1.09)	46.20
14. Feel "on the go", "driven by a motor"	398	1.37(1.09)	42.00	291	1.20(1.04)	35.80	106	1.83(1.09)	59.40
15. Talks excessively	400	1.32(1.06)	39.80	293	1.17(1.00)	34.10	106	1.74(1.10)	55.70
16. Blurts out answers before question	400	1.20(1.05)	35.80	293	1.07(0.98)	30.00	106	1.56(1.16)	51.90
17. Difficulty awaiting turn	400	1.21(1.01)	33.00	293	1.02(0.91)	24.60	106	1.73(1.08)	56.60
18. Interrupts/intrudes on others	398	1.14(0.96)	33.00	291	1.00(0.90)	27.00	106	1.56(0.99)	50.00

Table 4. Descriptive information of item level data of BAARS-C (Mean, Standard Deviation, % significantly endorsed).

	Full			Valid Only			Suspect Only		
	N	M(SD)	% endorsed	N	M(SD)	% endorsed	N	M(SD)	% endorsed
Inattention Symptoms									
1. Careless mistakes at work	396	1.63(0.91)	53.00	289	1.55(0.90)	48.80	106	1.88(0.91)	74.50
2. Poor sustaining attention for task	395	1.47(0.92)	45.50	289	1.35(0.87)	40.60	105	1.80(0.95)	68.90
3. Doesn't listen when spoken to	397	1.24(0.94)	35.50	291	1.09(0.88)	29.40	106	1.64(0.98)	53.80
4. Doesn't follow instructions, finish work	396	1.37(1.04)	40.50	290	1.21(1.02)	34.80	106	1.82(0.98)	64.20
5. Difficulty organizing tasks/activities	397	1.73(0.96)	56.30	290	1.63(0.93)	51.20	106	2.01(0.97)	71.70
6. Avoids tasks involving sustained effort	397	1.64(1.02)	55.00	291	1.49(1.00)	48.80	106	2.05(0.97)	79.20
7. Loses things necessary for tasks	393	1.68(1.03)	53.50	288	1.57(1.03)	48.10	104	1.99(0.97)	66.00
8. Easily distracted	398	2.14(0.90)	75.50	291	2.04(0.92)	71.30	106	2.42(0.73)	96.20
9. Forgetful in daily activities	398	1.63(0.96)	51.00	291	1.54(0.97)	45.70	106	1.91(0.90)	72.60
Hyperactivity/Impulsivity Symptoms									
10. Fidgets with hands/feet, squirms	397	2.00(0.98)	69.00	292	1.91(1.01)	64.80	105	2.27(0.82)	81.10
11. Leaves seat when seating is expected	397	0.97(1.07)	28.00	290	0.80(0.97)	22.50	106	1.42(1.22)	35.80
12. Feels restless	396	1.69(0.99)	56.80	290	1.56(0.98)	51.20	106	2.08(0.94)	83.00
13. Difficulties with leisure activities	398	1.11(1.02)	30.50	292	0.94(0.96)	23.50	106	1.56(1.02)	46.20
14. Feel "on the go", "driven by a motor”	395	1.43(1.11)	44.50	289	1.27(1.10)	39.90	105	1.87(1.00)	59.40
15. Talks excessively	397	1.49(1.16)	46.30	291	1.36(1.16)	41.30	106	1.85(1.13)	55.70
16. Blurts out answers before question	398	1.51(1.09)	48.00	291	1.41(1.06)	43.70	106	1.81(1.11)	51.90
17. Difficulty awaiting turn	397	1.44(1.03)	43.50	290	1.26(0.97)	35.50	106	1.94(1.02)	56.60
18. Interrupts/intrudes on others	398	1.29(1.04)	39.30	291	1.15(0.98)	33.80	106	1.70(1.08)	50.00

Table 5. Confirmatory Factor Loadings for CSS

	Full		*Valid Only*	
Inattention Symptoms	IA	H/I	IA	H/I
1. Careless mistakes at work	0.67	--	0.63	--
2. Poor sustaining attention for task	0.63	--	0.62	--
3. Doesn't listen when spoken to	0.68	--	0.61	--
4. Doesn't follow instructions, finish work	0.65	--	0.62	--
5. Difficulty organizing tasks/activities	0.62	--	0.62	--
6. Avoids tasks involving sustained effort	0.54	--	0.52	--
7. Loses things necessary for tasks	0.64	--	0.63	--
8. Easily distracted	0.71	--	0.68	--
9. Forgetful in daily activities	0.71	--	0.72	--
Hyperactivity/Impulsivity Symptoms				
10. Fidgets with hands/feet, squirms	--	0.62	--	0.64
11. Leaves seat when seating is expected	--	0.61	--	0.54
12. Feels restless	--	0.71	--	0.74
13. Difficulties with leisure activities	--	0.71	--	0.61
14. Feel "on the go", "driven by a motor"	--	0.61	--	0.57
15. Talks excessively	--	0.60	--	0.56
16. Blurts out answers before question	--	0.73	--	0.7
17. Difficulty awaiting turn	--	0.76	--	0.73
18. Interrupts/intrudes on others	--	0.73	--	0.66
	IA-H/I r = 0.62		IA-H/I r = 0.54	

IA= Inattention, H/I= Hyperactivity/Impulsivity

Table 6. Confirmatory Factor Loadings for BAARS-C

	Full		*Valid Only*	
Inattention	IA	H/I	IA	H/I
1. Careless mistakes at work	0.67	--	0.68	--
2. Poor sustaining attention for task	0.76	--	0.74	--
3. Doesn't listen when spoken to	0.74	--	0.70	--
4. Doesn't follow instructions, finish work	0.75	--	0.74	--
5. Difficulty organizing tasks/activities	0.67	--	0.69	--
6. Avoids tasks involving sustained effort	0.67	--	0.63	--
7. Loses things necessary for tasks	0.67	--	0.72	--
8. Easily distracted	0.83	--	0.82	--
9. Forgetful in daily activities	0.75	--	0.79	--
Hyperactivity/Impulsivity				
10. Fidgets with hands/feet, squirms	--	0.70	--	0.73
11. Leaves seat when seating is expected	--	0.75	--	0.71
12. Feels restless	--	0.72	--	0.73
13. Difficulties with leisure activities	--	0.70	--	0.68
14. Feel "on the go", "driven by a motor"	--	0.64	--	0.60
15. Talks excessively	--	0.65	--	0.63
16. Blurts out answers before question	--	0.76	--	0.78
17. Difficulty awaiting turn	--	0.81	--	0.82
18. Interrupts/intrudes on others	--	0.77	--	0.74
	IA-H/I r = 0.62		IA-H/I r = 0.64	

IA= Inattention, H/I= Hyperactivity/Impulsivity

Table 7. CSS-Full Sample IRT Parameters from the GRM for Inattention and Hyperactivity/Impulsivity Items

	Item Parameter Estimates						
	α	**β_1**	**s.e.**	**β_2**	**s.e.**	**β_3**	**s.e.**
Inattention Symptoms							
1. Careless mistakes at work	1.74	-2.17	0.19	-0.25	0.08	1.13	0.13
2. Poor sustaining attention for task	1.29	-2.46	0.26	-0.48	0.11	1.15	0.15
3. Doesn't listen when spoken to	1.30	-1.45	0.16	0.60	0.12	2.20	0.24
4. Doesn't follow instructions, finish work	1.61	-1.48	0.14	-0.03	0.09	1.04	0.13
5. Difficulty organizing tasks/activities	1.53	-2.16	0.20	-0.62	0.10	0.68	0.11
6. Avoids tasks involving sustained effort	1.08	-2.99	0.36	-0.97	0.15	0.56	0.13
7. Loses things necessary for tasks	1.44	-1.70	0.17	-0.16	0.09	0.95	0.13
8. Easily distracted	1.41	-4.13	0.57	-2.05	0.22	-0.51	0.10
9. Forgetful in daily activities	2.18	-2.16	0.18	-0.50	0.08	0.59	0.09
Hyperactivity/Impulsivity Symptoms							
10. Fidgets with hands/feet, squirms	1.19	-2.27	0.27	-0.99	0.15	0.25	0.11
11. Leaves seat when seating is expected	1.32	0.06	0.10	1.38	0.16	2.49	0.27
12. Feels restless	1.51	-2.27	0.23	-0.60	0.11	0.75	0.11
13. Difficulty with leisure activities	1.48	-0.75	0.11	0.80	0.11	1.72	0.17
14. Feel "on the go", "driven by a motor"	1.35	-0.99	0.14	0.31	0.10	1.29	0.15
15. Talks excessively	1.51	-0.95	0.12	0.36	0.09	1.34	0.14
16. Blurts out answers before question	1.81	-0.65	0.09	0.49	0.09	1.39	0.14
17. Difficulty awaiting turn	1.95	-0.79	0.10	0.58	0.09	1.37	0.12
18. Interrupts/intrudes on others	1.81	-0.76	0.10	0.59	0.10	1.75	0.17

Note. α = item discriminations, β_1 (Endorsement of 0 vs 1, 2, 3), β_2 (Endorsement of 0, 1, vs 2, 3), β_3 (0, 1, 2 vs. 3) = threshold categories, s.e.= standard error.

Table 8. Valid-only CSS IRT Parameters from the GRM for Inattention and Hyperactivity/Impulsivity Items

	Item Parameter Estimates						
	α	**β_1**	**s.e.**	**β_2**	**s.e.**	**β_3**	**s.e.**
Inattention Symptoms							
1. Careless mistakes at work	1.61	-2.08	0.22	-0.02	0.10	1.48	0.18
2. Poor sustaining attention for task	1.22	-2.45	0.31	-0.36	0.12	1.45	0.21
3. Doesn't listen when spoken to	1.15	-1.34	0.19	0.99	0.17	2.80	0.39
4. Doesn't follow instructions, finish work	1.54	-1.34	0.16	0.16	0.11	1.30	0.17
5. Difficulty organizing tasks/activities	1.47	-2.12	0.24	-0.55	0.11	0.87	0.14
6. Avoids tasks involving sustained effort	0.99	-2.94	0.43	-0.85	0.17	0.87	0.19
7. Loses things necessary for tasks	1.38	-1.60	0.19	0.01	0.11	1.16	0.17
8. Easily distracted	1.27	-4.23	0.67	-1.99	0.26	-0.24	0.12
9. Forgetful in daily activities	2.28	-2.09	0.20	-0.39	0.09	0.76	0.11
Hyperactivity/Impulsivity Symptoms							
10. Fidgets with hands/feet, squirms	1.15	-2.10	0.30	-0.84	0.16	0.47	0.13
11. Leaves seat when seating is expected	1.08	0.36	0.14	2.05	0.30	3.78	0.60
12. Feels restless	1.49	-2.23	0.27	-0.35	0.12	1.01	0.14
13. Difficulty with leisure activities	1.12	-0.62	0.16	1.33	0.20	2.76	0.38
14. Feel "on the go", "driven by a motor	1.18	-0.85	0.17	0.64	0.14	1.84	0.25
15. Talks excessively	1.33	-0.82	0.15	0.67	0.13	1.86	0.22
16. Blurts out answers before question	1.70	-0.57	0.12	0.75	0.11	1.78	0.19
17. Difficulty awaiting turn	1.82	-0.65	0.12	0.93	0.11	1.84	0.19
18. Interrupts/intrudes on others	1.66	-0.58	0.12	0.87	0.12	2.21	0.24

Note. α = item discriminations, β_1 (Endorsement of 0 vs 1, 2, 3), β_2 (Endorsement of 0, 1, vs 2, 3), β_3 (0, 1, 2 vs. 3) = threshold categories, s.e.= standard error.

Table 9. BAARS-C Full Sample IRT Parameters From the GRM for Inattention and Hyperactivity/Impulsivity Items

	Item Parameter Estimates						
	α	**β_1**	**s.e.**	**β_2**	**s.e.**	**β_3**	**s.e.**
Inattention Symptoms							
1. Careless mistakes at work	1.83	-1.79	1.60	-0.11	0.08	1.14	0.12
2. Poor sustaining attention for task	1.72	-1.51	0.14	0.13	0.09	1.43	0.14
3. Doesn't listen when spoken to	1.58	-1.04	0.12	0.55	0.10	1.75	0.17
4. Doesn't follow instructions, finish work	2.50	-0.86	0.09	0.31	0.08	1.03	0.10
5. Difficulty organizing tasks/activities	1.67	-1.86	0.17	-0.21	0.09	0.92	0.11
6. Avoids tasks involving sustained effort	1.58	-1.48	0.14	-0.20	0.09	0.99	0.12
7. Loses things necessary for tasks	1.55	-1.57	0.17	-0.12	0.09	0.89	0.11
8. Easily distracted	1.98	-2.25	0.20	-0.90	0.11	0.26	0.08
9. Forgetful in daily activities	2.17	-1.56	0.14	0.00	0.08	0.93	0.09
Hyperactivity/Impulsivity Symptoms							
10. Fidgets with hands/feet, squirms	1.37	-2.19	0.22	-0.79	0.12	0.45	0.11
11. Leaves seat when seating is expected	1.73	-0.17	0.09	0.80	0.11	1.53	0.16
12. Feels restless	1.55	-1.67	0.16	-0.27	0.09	0.99	0.13
13. Difficulty with leisure activities	1.39	-0.68	0.11	0.79	0.12	1.77	0.19
14. Feel "on the go", "driven by a motor	1.50	-0.98	0.12	0.20	0.09	1.10	0.14
15. Talks excessively	1.65	-0.91	0.11	0.12	0.09	0.81	0.12
16. Blurts out answers before question	2.09	-1.02	0.10	0.06	0.08	0.88	0.11
17. Difficulty awaiting turn	2.35	-1.03	0.09	0.21	0.08	1.03	0.11
18. Interrupts/intrudes on others	2.20	-0.79	0.09	0.35	0.08	1.27	0.13

Note. α = item discriminations, β_1 (Endorsement of 0 vs 1, 2, 3), β_2 (Endorsement of 0, 1, vs 2, 3), β_3 (0, 1, 2 vs. 3) = threshold categories, s.e.= standard error.

Table 10. Valid-only BAARS-C IRT Parameters from the GRM for Inattention and Hyperactivity/Impulsivity Items

	Item Parameter Estimates						
	α	**β_1**	**s.e.**	**β_2**	**s.e.**	**β_3**	**s.e.**
Inattention Symptoms							
1. Careless mistakes at work	1.81	-1.70	0.16	0.05	0.09	1.34	0.15
2. Poor sustaining attention for task	1.49	-1.50	0.17	0.33	0.10	1.91	0.22
3. Doesn't listen when spoken to	1.45	-0.91	0.13	0.84	0.13	2.27	0.26
4. Doesn't follow instructions, finish work	2.45	-0.66	0.08	0.5	0.08	1.28	0.12
5. Difficulty organizing tasks/activities	1.75	-1.76	0.17	-0.03	0.09	1.1	0.13
6. Avoids tasks involving sustained effort	1.38	-1.42	0.17	0.02	0.10	1.37	0.17
7. Loses things necessary for tasks	1.73	-1.37	0.15	0.09	0.09	1.01	0.13
8. Easily distracted	1.75	-2.18	0.21	-0.78	0.10	0.43	0.10
9. Forgetful in daily activities	2.51	-1.36	0.12	0.17	0.08	0.99	0.11
Hyperactivity/Impulsivity Symptoms							
10. Fidgets with hands/feet, squirms	1.39	-1.96	0.25	-0.60	0.14	0.56	0.12
11. Leaves seat when seating is expected	1.56	0.01	0.11	1.12	0.14	2.12	0.23
12. Feels restless	1.56	-1.51	0.19	-0.05	0.11	1.31	0.15
13. Difficulty with leisure activities	1.34	-0.44	0.14	1.16	0.16	2.12	0.25
14. Feel "on the go", "driven by a motor	1.37	-0.73	0.15	0.39	0.12	1.42	0.18
15. Talks excessively	1.54	-0.78	0.14	0.31	0.11	1.03	0.13
16. Blurts out answers before question	2.17	-0.92	0.14	0.21	0.09	1.07	0.11
17. Difficulty awaiting turn	2.44	-0.88	0.13	0.48	0.09	1.35	0.12
18. Interrupts/intrudes on others	2.09	-0.66	0.13	0.56	0.09	1.62	0.15

Note. α = item discriminations, β_1 (Endorsement of 0 vs 1, 2, 3), β_2 (Endorsement of 0, 1, vs 2, 3), β_3 (0, 1, 2 vs. 3) = threshold categories, s.e.= standard error.