The Psychometric Properties of the Barkley Adult ADHD Rating Scale—IV (BAARS-IV) in a College Sample

Rebecca Lynch
THE PSYCHOMETRIC PROPERTIES OF THE BARKLEY ADULT ADHD RATING SCALE—IV (BAARS-IV) IN A COLLEGE SAMPLE

By

REBECCA LYNCH

A Dissertation submitted to the Department of Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy

2017
Rebecca Lynch defended this dissertation on November 8, 2017.
The members of the supervisory committee were:

Janet A. Kistner
Professor Directing Dissertation

Insu Paek
University Representative

Sara Hart
Committee Member

Michael Kofler
Committee Member

Christopher Schatschneider
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.
# TABLE OF CONTENTS

List of Tables ....................................................................................................................................................... iv
List of Figures ......................................................................................................................................................... v
Abstract ................................................................................................................................................................... vi

1. INTRODUCTION ...................................................................................................................................................... 1

2. METHODS ................................................................................................................................................................. 21

3. RESULTS .................................................................................................................................................................. 29

4. DISCUSSION ............................................................................................................................................................. 37

APPENDICES ............................................................................................................................................................... 66

A. INSTITUTIONAL REVIEW BOARD PROJECT APPROVAL MEMORANDUM .............. 66
B. INSTITUTIONAL REVIEW BOARD PROJECT RENEWAL MEMORANDUM ............ 68

References ................................................................................................................................................................. 70

Biographical Sketch ....................................................................................................................................................... 84
LIST OF TABLES

1  Demographic Characteristics of the Study Sample .................................................................50
2  Means and Standard Deviations for the BAARS-IV Scales Raw Scores for the Total Sample and Gender Separately ..................................................................................................................51
3  Model Fit Statistics for Alternative Models of ADHD Symptoms ......................................52
4  Item Parameters for the BAARS-IV in the Logistic Metric ......................................................53
5  Discrimination Parameters and Threshold Values Based on Gender, Significance Levels, and Effect Sizes of DIF ...............................................................................................................................54
# LIST OF FIGURES

1. Schematic representations of the two correlated models of ADHD symptoms ......................55  
2. Schematic representations of the two bifactor models of ADHD symptoms ....................56  
3. Correlated factor models of ADHD symptoms in college students ................................57  
4. Two-factor bifactor model of ADHD symptoms in college students .............................58  
5. Item information functions for inattention items (Panel A), and the test information function (Panel B) and its standard error (Panel C) for the Inattention (IA) scale .............................59  
6. Category response curves for inattention items ..................................................................60  
7. Item information functions for hyperactivity items (Panel A), and the test information function (Panel B) and its standard error (Panel C) for the Hyperactivity (HYP) scale ..........62  
8. Category response curves for hyperactivity items .............................................................63  
9. Item information functions for impulsivity items (Panel A), and the test information function (Panel B) and its standard error (Panel C) for the Impulsivity (IMP) scale .......................64  
10. Category response curves for impulsivity items ..............................................................65
ABSTRACT

Attention-Deficit/Hyperactivity Disorder (ADHD) is a chronic disorder associated with long-term impairment across multiple life domains, including academic, occupational, social, and psychological (Barkley, Murphy, & Kwasnik, 1996; Harpin, 2005; Wilens, Biederman, & Spencer, 2002). Increasing numbers of college students are presenting to health centers and counseling programs with complaints of inattention, distractibility, and restlessness, underscoring the critical need for reliable and valid methods for evaluating ADHD in this age group (e.g., U.S. Government Accountability Office [GAO], 2009). There is a growing body of research suggesting that the Barkley Adult ADHD Rating Scale—IV (BAARS-IV; Barkley, 2011) is a reliable tool for assessing ADHD in adults (e.g., Becker, Marshall, & McBurnett, 2014), yet there are a number of limitations and omissions in existing data. The current study investigated the psychometric properties of the BAARS-IV in a clinic-referred sample of 607 college students between the ages of 18 and 25. First, the factor structure of ADHD was examined by comparing traditional two-factor and three-factor models of ADHD with bifactor models of ADHD using confirmatory factor analyses (CFA). The correlated three-factor model consisting of separate dimensions of inattention, hyperactivity, and impulsivity was considered the best representation of the ADHD constructs for this sample. Next, a multidimensional graded response model (GRM) based on item response theory (IRT) was applied to the data. The results of the GRM indicated that most items of the BAARS-IV showed adequate discrimination of their respective latent traits (i.e., Inattention, Hyperactivity, and Impulsivity) and functioned best when measuring participants with average levels of the latent traits. Of the 18 items, only seven items measure sub-clinical levels of their respective latent trait; however, two of these items provided relatively less information. The remaining five items (one hyperactivity and four impulsivity) had good discrimination and difficulty parameters, providing useful information at sub-clinical levels of the latent traits. The final aim of this study was to examine the items for potential differential item functioning (DIF) by gender and to estimate the size and impacts of detected DIF. The results revealed significant DIF for two inattention items, two hyperactivity items, and one impulsivity item. The findings from the current study suggest that some items of the BAARS-IV may have more clinical utility than the others in the assessment of ADHD in college students.
CHAPTER 1

INTRODUCTION

Background

Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder characterized by inattention, hyperactivity, and impulsivity and is one of the most prevalent mental health disorders among children, with prevalence rates in the United States being between 3% and 11% (Akinbami, Liu, Pastor, & Reuben, 2011; American Psychiatric Association [APA], 2013; Goldman, Genel, Bexman, & Slanetz, 1998; Visser et al., 2014). According to the APA, ADHD is more common in boys than girls, with a ratio of 2:1 (APA, 2013). The disorder accounts for 30% to 50% of child referrals to mental health services and is associated with a variety of deficits and impairments in cognitive, social, family, and school functioning (Barkley, Fischer, Edelbrock, & Smallish, 1990; Biederman et al., 1996; Faraone & Biederman, 1998). The problems associated with ADHD place children at increased risk for accidental injury, peer rejection, driving accidents, and school dropout (Biederman et al., 2006; Mikami & Hinshaw, 2006; Rapport, Scanlan, & Denney, 1999; Swenson et al., 2004).

ADHD has been shown to be a chronic disorder, with most elementary school-age children continuing to exhibit significant symptoms into and beyond middle and high school. One longitudinal study followed a sample of 94 children (5−11 years old) with ADHD for eight years and found that 44% of the children continued to exhibit clinically significant symptoms in adolescence (Bussing, Mason, Bell, Porter, & Garvan, 2010); however, a review of the persistence of ADHD conducted by Faraone and colleagues (2006) suggested the rate of persistence is higher. Findings from a large nationally representative sample showed that roughly 6.5% of 13- to 17-year-olds met diagnostic criteria for ADHD based on the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM-IV; 1996; American Psychiatric Association [APA]; Kessler et al. 2012). Moreover, longitudinal studies have documented that the majority of children and adolescents diagnosed with ADHD display symptoms into adulthood, with 30% to 80% of children diagnosed with ADHD continuing to exhibit impairment associated with the disorder as adults (e.g., Biederman et al., 1996; Biederman, Petty, Evans, Small, & Faraone, 2010; Weiss & Hechtman, 1993). Prevalence estimates of ADHD in the adult
population ranges from 2% to 5% (APA, 2013; Kessler et al., 2006), meaning that at least 8 million adults in the US have the disorder.

There is a substantial quantity of research available concerning ADHD in children and adolescents, but relatively little is known about ADHD in adults. In particular, there is a lack of knowledge regarding ADHD in young adults, such as college students. As a group, high school students with ADHD are less likely than their peers to graduate from high school and attend college (DuPaul, Weyandt, O’Dell, & Varejao, 2009). Nevertheless, surveys have found that 2% to 9% of the college population meets diagnostic criteria for ADHD (DuPaul et al., 2009; Pryor, Hurtado, DeAngelo, Palucki Blake, & Tran, 2010), though the actual percentage of college students with ADHD is thought to be much higher given that students are not required to report their disability to colleges and universities (Garnier-Dykstra, Pinchevsky, Caldeira, Vincent, & Arria, 2010; Pryor et al., 2010). Moreover, this number is on the rise, largely due to a growing number of high school students with ADHD that are graduating and then attending college, but also because there has been a significant increase in the number of college students presenting to health centers with complaints of symptoms associated with ADHD (Weyandt & DuPaul, 2008; Wolf, Simkowitz, & Carlson, 2009). Consequently, many college-based (e.g., psychology clinic, counseling center) and community-based clinicians have witnessed dramatic increases in the number of students requesting evaluations and services based on a documented diagnosis of ADHD, leading to increased demands for clinicians to validate their disability and eligibility for specific accommodations (DuPaul et al., 2009; Wolf et al., 2009). Students with disabilities represent about 11% of postsecondary students; of that number, students with ADHD comprise approximately 19.1% (U.S. Government Accountability Office [GAO], 2009). This rate nearly tripled from 2000 to 2008, rising from 6.7% to 19.1%, making it the second most common disability, whereas the next most rapidly increasing disability category increased only 7.2% (i.e., mental, emotional, or psychiatric condition, or depression; U.S. GAO, 2009). Since then, it has been reported that 25% of college students receiving disability services through school were diagnosed with ADHD (DuPaul et al., 2009; Wolf et al., 2009).

The increased prevalence of ADHD on college campuses has led to an increase in the attention researchers have paid to the specific challenges and difficulties faced by students with ADHD. College students with ADHD are likely to experience a number of academic, psychological, and social problems (Weyandt & DuPaul, 2012; Weyandt et al., 2013). With
respect to academic functioning, undergraduate students with ADHD have been shown to have lower GPAs and graduation rates, higher rates of academic-related problems (e.g., academic probation), and lower “motivation,” study habits, and study skills (Advokat, Lane & Luo, 2011; Barkley, 2008; Heiligenstein, Conyers, Berns, & Miller, 1998; Norwalk, Norvilitis, & MacLean, 2009; Reaser, Prevatt, Petscher, & Proctor, 2007) than those without ADHD. Studies with college students have generally found that students with ADHD endorse higher levels of psychoticism, anger, aggression, depression, anxiety, and substance use (e.g., Ramirez et al., 1997; Richards, Rosén, & Ramirez 1999; Shaw-Zirt, Popali-Lehane, Chaplin, & Bergman, 2005; Theriault & Holmberg, 2001; Upadhyaya et al., 2005; Wymbs et al., 2012). College students with documented ADHD also report lower levels of life satisfaction and higher levels of driving-related problems (e.g., increased car accidents, drinking and driving, overall driving citations, and license suspensions and revocations; Barkley, Murphy, DuPaul, & Bush, 2002; Grenwald-Mayes, 2002; Woodward, Fergusson, & Horwood, 2000). The social problems of college students with ADHD are not well established, but a few studies have shown that ADHD symptoms are negatively associated with social skills, number of friends, and positively associated with anger, emotion dysregulation, and difficulty in recognizing emotion in others (Nijmeijer et al., 2008; Rapport, Friedman, Tzelepis, & Van Voorhis, 2002; Reimherr et al., 2007; Shaw-Zirt et al., 2005). Furthermore, young adults who do not receive treatment are at even greater risk for academic failure, work/relationship difficulties, and car accidents (Barkley et al., 2006; Barkley, Guevremont, Anastopoulos, DuPaul, & Shelton, 1993; Barkley et al., 1996; Harpin, Mazzone, Raynaud, Kahle, & Hodgkins, 2013; Heiligenstein & Keeling, 1995).

It is important to emphasize that the challenges faced by college students with ADHD are somewhat distinct from those found in young adults with ADHD in general. In addition to the developmental changes occurring as they transition into adulthood, college students must also adapt to new environments, learn to manage greater academic, social, and organizational demands, and deal with decreased parental support (Blanco et al., 2008; Hunt & Eisenberg, 2010). In this way, college students with ADHD represent a unique group of individuals suffering from the disorder. Given the chronicity of the disorder, the frequency with which it occurs among college-aged young adults, and the significant academic and/or social impairments college students with ADHD are likely to experience, it is crucial that valid and reliable methods
of identifying and diagnosing ADHD in this population are available so that they are eligible for accommodations and/or treatment.

Assessment of ADHD in College Students

The prevalence of ADHD on college campuses (i.e., 2-8%; DuPaul et al., 2009; Pryor et al., 2010) and federal law (the Americans With Disabilities Act [ADA]; 2008) mandating that postsecondary institutions provide reasonable services accommodations to “otherwise qualified” students with disabilities have both contributed to the significant increase in the number of college students requesting ADHD evaluations (Pryor et al., 2010). Furthermore, there has been an increase in the proliferation of an ADHD diagnosis among college students seeking accommodations or prescription medication for the purposes of cognitive enhancement or other types of misuse (e.g., Barrett, Darredeau, Bordy, & Pihl, 2005; Booksh, Pella, Singh, & Gouview, 2010; Harrison, 2004; Harrison, 2006; Hartung et al., 2013; McCabe, Knight, Teter, & Wechsler, 2005; Weyandt et al., 2014). Thus, it is important that assessments for ADHD in college students are comprehensive such that they include multiple methods and respondents. Specific components may include clinical interviews with the student and significant others (e.g., parents, roommates, significant others), self- and other-report ratings scales, intellectual and achievement testing, and review of the client’s records for information regarding academic functioning and previous diagnostic evaluations (e.g., Barkley, 2006; Barkley, 2011; Montano, 2004). Although comprehensive evaluations are necessary for establishing a diagnosis of ADHD in college students and for procuring support services, they can be quite costly and time consuming. As such, if a student is reporting academic, behavioral, and/or social difficulties at the time of referral, they may be asked to fill-out a brief rating scale regarding their current ADHD symptoms. Rating scales used in this respect offer an efficient and cost-effective method of obtaining information regarding the presence and severity of the student’s ADHD symptoms, and can be used to help the clinician decide whether a comprehensive diagnostic evaluation of ADHD is warranted (Barkley, 2011). To determine whether further diagnostic evaluation is needed, the clinician may compare the overall score with a cutoff score or check to see whether the ratings exceed a particular threshold (e.g., 90th percentile based on respondent’s sex and age) or use a symptom count cutoff score (e.g., six symptoms from either domain; Barkley, 2011; Smith, Jr. & Johnson, 2000).
**Adult ADHD Rating Scales**

Rating scales can be highly useful in screening students for possible ADHD symptoms because they provide a quantifiable measure of frequency, duration, and/or severity. However, rating scales are also often utilized for providing adjunctive diagnostic data and evaluating change in ADHD symptoms resulting from interventions (e.g., ADHD medications; Barkley, 2011; Taylor, Deb, & Unwin, 2011). In general, other advantages of ratings scales are ease of use, ability to capture behaviors that may not be observable in the clinic, ability to cover a large variety of items and constructs within a relatively short time, possibility of obtaining ratings from other informants, instrument administration normally does not require a specialized set of skills to administer, generally good to excellent test-retest reliability, and strong internal consistency (Barkley, 2011; Hinshaw & Nigg, 1999).

There are a number of well-validated ADHD scales available for use with children. In adults, however, the availability of well-validated ADHD measures is limited. A review of the literature to identify self-report scales that are most relevant to the clinical assessment of current ADHD symptomatology in adults yielded only four rating scales. Although there are several other instruments that are used in the assessment of adult ADHD, most are either limited by lack of proper testing and validation, or are designed for use in a longer, complex process of establishing or ruling-out comorbid internalizing and externalizing conditions. The four scales that were identified are those that were developed specifically for the assessment of adult symptoms of ADHD, include items that reflect all of the *DSM-5* (APA, 2013) symptoms of ADHD, are accessible to practicing clinicians (i.e., available either in the public domain or through a publishing company), and are available in English: (1) Adult Attention Deficit Disorders Evaluation Scale (A-ADDES; McCarney & Anderson, 1996); (2) the ADHD Self-Report Scale v1.1 Symptom Checklist (ASRS-V1.1; Kessler et al., 2005); (3) the Barkley Adult ADHD Rating Scale—IV (BAARS-IV: Barkley, 2011); and; (4) the Conners’ Adult ADHD Rating Scale (CAARS: Conners, Erhardt, and Sparrow, 1999). Whilst an in-depth discussion of these scales is beyond the scope of this paper, a review that includes some of these scales is available (Taylor et al., 2011). The focus of the current study is the BAARS-IV, as it is emerging as one of the most commonly used adult ADHD rating scales and has good psychometric properties.
The BAARS-IV is a recently developed tool for assessing current ADHD symptoms and domains of impairment as well as recollections of childhood symptoms in adults between the ages of 18 and 89 (Barkley, 2011). There are various versions of the BAARS-IV, but for the purposes of the present paper, the self-report version for current symptoms will be the focus of the remainder of this discussion. The self-report version for current symptoms is a 30-item scale comprised of 18 items that correspond to the *DSM-5* ADHD symptoms (nine items of inattention, five items of hyperactivity, and four items of impulsivity) as well as nine items evaluating sluggish cognitive tempo (SCT; e.g., easily confused, slow moving). The frequency and severity of each item is rated for the past six months on a 4-point Likert scale: (1) never or rarely, (2) sometimes, (3) often, and (4) very often. Information on age of onset and impairment domains is also collected. Responses to the 27 main items can yield nine scores: four subscale total raw scores (ADHD Inattention, ADHD Hyperactivity, ADHD Impulsivity, and SCT), an ADHD Total Score (sum of the three ADHD scores), and four ADHD symptom count scores (Inattention, Hyperactivity/Impulsivity, total ADHD, and SCT). The BAARS-IV takes approximately five to seven minutes to complete.

The BAARS-IV evolved from previous scales developed by its author and his colleagues (e.g., the Adult ADHD Rating Scale, Barkley’s Current Symptoms Scale, the Prototype BAARS [P-BAARS]; Barkley, 2008; Barkley & Murphy, 2005; Murphy & Barkley, 1996). In the development of the BAARS-IV, the item pool consisted of the 18 *DSM-IV-TR* (APA, 2000) symptoms along with a question concerning the estimated onset of symptoms and whether or not they resulted in impairment in 10 major functional domains. These impairment domains were reduced so that they were more in line with the *DSM-5*. Also in line with changes made for the current version of the *DSM*, the 18-items from the BAARS-IV were slightly modified in language to better fit adult symptoms of ADHD (e.g., references to school/schoolwork were removed, “play” activities was replaced with “fun”). Also new to the BAARS-IV was the addition of nine items for evaluating SCT, which is believed to characterize a subset of adults who are often diagnosed with inattentive type (Barkley, 2011).

The BAARS-IV was normed on a U.S. sample of 1,249 adults ranging in age from 18 to 89. Regarding the composition of the normative sample, it was comparable to the 2000 U.S. Census estimates in terms of gender, ethnicity, income, marital status, and employment status. From these data, adult norms were reported and cutoff criteria for clinical significance were
established. Based on the results of a factor analysis conducted on the 18 *DSM*-based items, the BAARS-IV was found to yield three separate dimensions of inattention, hyperactivity, and impulsivity (Barkley, 2011). Becker, Langberg, Luebbe, Dvorsky, and Flannery (2014) reported finding a similar factor structure in a large sample of college students. A similar pattern of results has been reported in other studies using an earlier version of the BAARS-IV. Specifically, this factor structure was replicated in a non-clinical sample of adults in the Netherlands (Kooij, Buitelaar, Van Den OORD, Furer, Th Rijnders, & Hodiamont, 2005), in a study of young adults (Caterino, Gomez-Benito, Balleurka, & Amador-Compos, 2009), and in a sample of college students (i.e., for both self and parent ratings using a prototype version of the BAARS-IV; Proctor & Prevatt, 2009).

According to the manual, the four ADHD scales demonstrate internal consistency coefficients ranging from satisfactory to excellent ($\alpha = .90$, Inattention; $\alpha = .78$, Hyperactivity; $\alpha = .81$, Impulsivity; $\alpha = .91$, ADHD Total Score). Recent studies have reported similar estimates when using the BAARS-IV (Flannery, Becker, & Luebbe, 2014; Voinescu, Szentagotai, & David, 2012). With respect to test-retest reliability, the subscales were reported to have adequate test-retest reliability over a 2- to 3- week period ($r = .66$, Inattention; $r = .72$, Hyperactivity; $r = .76$, Impulsivity; $r = .75$, ADHD Total Score; Barkley, 2011). As evidence for the convergent validity of this measure, the manual reports that the BAARS-IV subscales share a significant amount of their variance with executive functioning deficits as measured by ratings on the Barkley Deficits in Executive Functioning Scale (BDEFS). The majority of the remaining psychometric data on the BAARS-IV is based on studies using the earlier version of the BAARS-IV mentioned previously. Briefly, inter-rater reliability has been reported to be fair to excellent (Barkley, Knouse, & Murphy, 2011), significant correlations have been found between the P-BAARS and Conners’ Continuous Performance Test scores providing additional evidence of convergent validity (Barkley, Murphy, & Fischer, 2008), and in terms of criterion validity, the P-BAARS was found to correlate highly with a structured clinical interview (Barkley, 2011).

In terms of clinical utility, the BAARS-IV can be used to evaluate symptoms of ADHD quickly and inexpensively to screen individuals to help determine whether comprehensive clinical assessment is warranted or as part of a comprehensive assessment in diagnosing ADHD, research settings where the evaluation of ADHD symptoms is of interest, and for assessing treatment effects (Barkley, 2011). The authors of the scale suggest that the self- and other-report
forms for current and childhood symptoms, along with assessments of developmental history, employment, medical health, social history, and driving behaviors can be used in conjunction in performing a structured diagnostic ADHD assessment in adults (Barkley, 2011). Given some evidence to suggest that adults with ADHD tend to underestimate their own ADHD symptomatology (Manor et al., 2010), the availability of informant (collateral) versions of the BAARS-IV is a particularly noteworthy attribute. The corresponding manual is comprehensive and provides clear administration and scoring instructions. Currently, there is no online or software-based administration or scoring; however, there are multiple versions of the scale so there are opportunities to collect information from collateral sources. The manual, which also includes an interview version of the scale, is available through Guilford Press for $149 (https://www.guilford.com/books/Barkley-Adult-ADHD-Rating-Scale-IV-BAARS-IV-Russell-Barkley/9781609182038). Purchase of the manual carries with it permission to photocopy the scales, meaning there is no additional cost for the BAARS-IV forms.

**Future Directions and Gaps in the Literature**

**An Item Response Theory Approach**

The BAARS-IV has the potential to be a tool for health care professionals and researchers to use with adults to assess ADHD symptomatology. To date, the majority of studies examining the psychometric properties of the BAARS-IV have been derived from analyses based on the traditional classical test theory (CTT; Lord & Novick, 1968) approach. There is a lack of research that has analyzed the psychometric properties of the BAARS-IV using modern approaches. This gap warrants attention as the use of the BAARS-IV is becoming more common in clinical and research settings, and because of an urgent need for valid and age-appropriate tools for use with young adults. Moreover, given the direct relation between the BAARS-IV and the DSM-5 criteria for ADHD, further examination of the psychometric properties of the BAARS-IV has the potential to provide valuable insight into other important gaps in our understanding of ADHD in young adults. The purpose of the current study is to advance our knowledge of the psychometric properties of the BAARS-IV. To this end, the psychometric properties will be evaluated using both a CTT approach through confirmatory factor analyses as well as item response theory (IRT; Embretson & Reise, 2000) techniques.
**A Brief Overview of IRT and CTT**

Item response theory, also referred to as latent trait theory, is a model-based method of latent trait measurement that estimates the relationship between an individual’s responses to items and the ability or trait that each item is intended to measure (Embretson & Reise, 2000). IRT can be used not only to obtain information on the quality or precision of a scale as a whole but also for individual items as well as specific ranges of the instrument (e.g., around a specified cutoff point). While both CTT and IRT techniques provide information pertaining to the psychometric properties of a measure, there are a number of advantages of using IRT techniques (see Embretson & Reise, 2000 for a detailed discussion of the advantages of IRT relative to CTT). First, CTT properties (e.g., reliability, item-total correlation, and standard error of measurement) are sample dependent and, therefore, vary across samples. In contrast, IRT item parameters are not dependent on the sample used to generate the parameters, and are assumed to be group invariant. Second, CTT often assumes that all items carry the same amount of information about the construct being measured; whereas, IRT allows for the importance (or weight) of individual items to vary. Third, CTT produces only a single estimate of reliability and standard error of measurement (SEM) for an entire instrument, whereas IRT yields reliability and SE values for each item of a scale, and this is provided for all trait levels across the trait spectrum. Because IRT analysis can be used to yield item- and scale-level characteristics of the BAARS-IV, this study can contribute to the research on the difficulty and amount of reliable information that the items and subscales of the BAARS-IV provide.

**IRT Parameters and Information**

In IRT-based analysis, parameters for both an individual’s latent trait and symptom (item) functioning are estimated. In IRT, a *theta* ($\theta$) value reflects the individual’s position on the latent trait continuum (Embretson & Reise, 2000). A number of IRT models can be employed to try and provide an estimate of the probability of an item being endorsed using information about the person being rated and the item, but the most commonly used are one-, two-, and three-parameter models. In the two-parameter logistic (2PL) model, the model yields a discrimination parameter or slope ($a$) and a difficulty parameter or threshold ($b$). The discrimination parameter for an item corresponds to a factor loading and provides an estimate of how well an item distinguishes between individuals at the thresholds (i.e., high vs. low ADHD). In general, higher
discrimination parameters are considered better, as they tend to provide more information. For example, a low value would mean that the probability of endorsing that item is about the same at high and low levels of the latent trait. As such, the discrimination parameter can be useful for selecting items that differentiate well between individuals with low and high levels of the trait as measured by test items. The difficulty parameter signifies the level of the latent trait needed for a response to be endorsed.

IRT models are categorized based on the response formats of the items that comprise the measure of interest. Items that have only two responses (e.g., yes/no) are referred to as dichotomous, while items with multiple ordered-response categories are referred to as polychotomous. For measures with polychotomous item response data, as is the case with the BAARS-IV, a polychotomous IRT model is needed to represent the relationship between an individual’s trait level and the probability of endorsing a particular response option. One commonly used polychotomous IRT model is the graded-response model (GRM; Samejima, 1969; Samejima, 1997), which is considered an extension of the 2PL model and was specifically designed for use with Likert scales. With a GRM, one discrimination parameter is generated for each item and one difficulty parameter is generated for each between category threshold. For example, because the BAARS-IV has a 4-point Likert scale, three b parameters (b1, b2, and b3) will be estimated. These parameters represent points along the latent trait continuum at which there is a 50% chance of endorsing the item at that threshold (for a polychotomous scale). For example, the first threshold parameter for each item represents the amount of overall behavioral problems needed for an individual to have a 50% chance of endorsing a rating of one (not at all) and 50% chance of endorsing a rating of two (sometimes) or greater.

As previously stated, prior work based on CTT has indicated that the BAARS-IV has adequate to excellent reliability. Exploring the reliability of this scale within an IRT framework, however, can also generate indices that convey the amount of information that the test (or subscale) provides across varying levels of the latent trait (test information function; TIF), as well as show the amount of information obtained from individual items across the latent trait (item information functions; IIFs). The TIF of a scale is equivalent to the combined value of the IIFs for that scale (Reise, Ainsworth, & Haviland, 2005). The IIFs and TIFs are graphed using the discrimination and difficulty parameters, where greater information, as determined by the discrimination parameter estimates, indicates greater precision for measuring theta, and the
difficulty parameter estimates determine the location of the information curve across the range of theta. The standard error of measurement (SEM) is equivalent to the inverse of the square root of the TIF and is related to the reliability of the measurement. Together, the TIF and SEM represent the information and precision of measurement of a group of items across the spectrum of behaviors.

Based on a review of the literature, only three other studies have used IRT-based analyses to investigate the psychometric properties of DSM-based ADHD rating scales in adult samples (Gomez, 2011; Sanchez-Garcia et al., 2015; Smith, Jr. & Johnson, 2000). Gomez (2011) used a GRM to evaluate the psychometric properties of the Current Symptoms Scale (CSS; Barkley & Murphy, 1998) in a sample of Australian adults from the general community. The results showed that most of the symptoms were good at discriminating different levels of the latent traits (i.e., inattention, hyperactivity, and impulsivity). However, two symptoms of inattention (“don’t listen and “loses things”) and one symptom of impulsivity (“talks excessively”) were found to have poor discrimination ability when compared with the other symptoms on their respective traits. Further, low IIFs for “don’t listen,” “loses things,” and “driven by a motor” across the latent trait indicated that these symptoms did not measure their traits with adequate accuracy. With respect to scale-level information, all three scales showed adequate test information values (reliability) of the latent traits at the mean level onward, but were not reliable for individuals with relatively low levels of ADHD symptoms. This study did not, however, test competing models of the structure of ADHD (i.e. traditional correlated factor models and bifactor models) or test the ADHD symptoms for gender equivalency. Also, the Gomez (2011) study utilized the CSS, which has slightly different item wording than the BAARS-IV, so it is possible that the findings will not generalize to the BAARS-IV. Furthermore, the sample included adults between the ages of 18 and 50 years old, so the findings may not generalize to all adults (e.g., young adults who are in college).

The other two studies that have conducted IRT-based analyses to examine the psychometric properties of DSM-based ADHD rating scales in adult samples used a Rasch model (Sanchez-Garcia et al., 2015; Smith, Jr. & Johnson, 2000). Although Rasch models have certain benefits over other models, they assume that all items in a measure have the same discrimination value so the parameter estimates only include a difficulty parameter. Furthermore, Smith & Johnson’s (2000) study included a modified measure of ADHD that was not developed for use in
adults and the analysis did not include all 18 DSM symptoms. The sample used in the study conducted by Sanchez-Garcia and colleagues (2015) was entirely comprised of Spanish adults with substance use disorders, and the majority were men. Thus, these two studies may be related to the current study, but the findings may not be particularly informative. Considering the paucity of studies using an IRT approach to examine the psychometric properties of adult ADHD ratings scales in the general literature, there is a need for further research in this area. This gap in the literature is further complicated by concerns regarding the DSM diagnostic criteria for ADHD. A general issue in the ADHD literature is the validity of the diagnostic criteria for ADHD in adults.

The current diagnostic criteria of the DSM-5 are largely consistent with the previous criteria (DSM-IV-TR), which were originally created based on numerous studies and field trials on children and adolescents 4 to 17 years of age (Lahey et al., 1994; Spitzer, Davies, & Barkley, 1990). As such, the DSM-IV symptom criteria for ADHD were developed with children in mind and are not appropriately worded for adults and may not necessarily account for variation in the adult manifestation of ADHD. A noted change in the DSM-5 is that the some of the symptom criteria have been slightly reworded and include new symptom descriptors intended for adolescents and adults. For example, “often leaves seat in classroom or in other situations in which remaining seated is expected” was revised to “often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplace, or in other situations that require remaining in place).” Despite efforts to modify the symptoms so that they are age-appropriate adults, questions regarding validity of the symptoms remain (e.g., Adler & Cohen, 2004; Proctor & Prevatt, 2009). Symptoms measuring hyperactivity and impulsivity, in particular, appear age-limited. For example, some of the symptoms (e.g., “often unable to play or engage in leisure activities quietly”) still seem less relevant for adults than they are for children. Relatedly, symptoms of hyperactivity and impulsivity have been shown to decrease the most with increasing age, suggesting that they may not adequately capture the manifestation of these behaviors in adults (Barkley, 2006; Millstein, Wilens, Biederman, & Spencer, 1997). Further, these symptoms still have not been empirically validated with adults, including college students. Consequently, measures based on the DSM criteria may have fairly limited utility for the identification and diagnosis of ADHD in adults (Barkley et al., 2008). By further investigating the psychometric properties of the BAARS-IV,
this study has the potential to provide some insight with respect to the relative utility of each individual ADHD item.

**Dimensionality**

Before the psychometric properties of the BAARS-IV can be examined within an IRT framework, it is necessary to first establish unidimensionality of a scale. Unidimensionality is one of the major assumptions of IRT analyses and refers to the notion that a set of items measures a single underlying trait (Lord & Novick, 1968). It is important to explore dimensionality of the data prior to application of an IRT model because if parameters obtained with a unidimensional model do not accurately reflect the relations among items and the general latent factor, the parameters will likely be distorted in some way. As a result, IIFs and trait level estimates will be inaccurate, and the DIF analyses will be questionable. That said, the assumption of unidimensionality is rarely met in a strict sense because most measures of personality and psychopathology are fundamentally multidimensional, in that there is a sampling of items from multiple domains of an overall psychological construct (Reise & Revicki, 2014). Procedures have been developed for determining whether the data are “unidimensional enough,” but there are also multidimensional IRT models that can be fit to the data (Ackerman, Gierl, & Walker, 2003).

Beyond these psychometric objectives, exploring the dimensionality of item responses derived from the BAARS-IV is of substantive interest. Indeed, an area of ongoing debate in the ADHD literature is the structural organization of ADHD symptoms. This is apparent in the DSM, as the most prominent changes to the DSM diagnostic criteria for ADHD have involved different organizations of inattention, hyperactivity, and impulsivity and the corresponding implications those changes have had for diagnosing subtypes. Specifically, ADHD was conceptualized as three factors (inattention, hyperactivity, and impulsivity) in the DSM-III (APA, 1980), one general factor in the DSM-III-R (APA, 1987), and two factors (inattention and hyperactivity/impulsivity) in the DSM-IV (APA, 1994). In the current version of the DSM, there has been a shift away from using subtypes and instead three symptom presentations are considered (i.e., inattentive, hyperactive/impulsive, and combined presentation). This change is rather significant given that it suggests that ADHD is best represented as a single disorder without subtypes.
It also ties into a growing body of research that has found support for a bifactor model, also referred to as a general factor model, of ADHD symptoms as opposed to the traditional one-, two-, and three-factor models for ADHD symptoms. Briefly, when applied to ADHD, a bifactor model includes a general (overall) factor of ADHD on which all symptoms load, as well as two (or three) specific factors on which the inattentive and hyperactive/impulsive (or inattentive, hyperactive, and impulsive) symptoms also load. An assumption of these bifactor models is that the general factor accounts for common variance in all of the 18 ADHD items that is independent of the specific factors (e.g., inattention, hyperactivity and/or impulsivity), whereas the specific factors account for variance in their respective symptom sets that is independent of the general factor. For example, an individual’s response to the item “how often do you have trouble wrapping up the final details of a project, once the challenging parts have been done?” is considered a function of their severity of ADHD as well as an indication of their level of inattention. These general and specific factors must be orthogonal (independent) for empirical identification of the bifactor model to occur. Thus, they are different than correlated factor models, which include a general higher-order factor that accounts for common variance across lower-order oblique factors and common item variance is attributed to the intercorrelations between factors.

At least six studies have examined bifactor models of ADHD in adults using ADHD rating scales (Gibbins, Toplak, Flora, Weiss, & Tannock, 2012; Gomez, Kyriakides, & Devlin, 2014; Gomez, Vance, & Gomez, 2013; Martel, Von Eye, & Nigg, 2012; Morin, Tran, & Caci, 2013; Sorge et al., 2015). In a sample of 751 adults with ADHD, Gibbins et al. (2012) found that a bifactor factor structure with one ADHD general factor and three specific factors fit the data better than a hierarchical model with two specific factors. Notably, the three-factor bifactor model of ADHD symptoms reported in Gibbins et al. (2012) corresponded more closely to the hyperactivity and impulsivity symptom groups used for hyperkinetic disorder (HKD) in the *International Classification of Diseases-10 (ICD-10)*; World Health Organization [WHO], 1993) than the *DSM-5* symptom groups such that item 15 “talks excessively” was allocated to the Impulsivity factor rather than the Hyperactivity factor. A similar pattern of results was replicated in one other study such that a general factor model with three specific factors (inattention, hyperactivity, and impulsivity) best represented the structure of ADHD in a community samples of adults living in France (Morin et al., 2013). Sorge and colleagues (2015) also reported that a
three-factor bifactor model provided a better fit to the data than alternative models in a population-based sample of Canadian adults; however, because this study relied on a 10-item scale of ADHD that was developed by the authors for their study, the conclusions that can be drawn regarding the current characterization of the DSM-5 ADHD symptoms are limited. Findings from the other available studies that have tested bifactor models of ADHD among adults were more in line with those from studies conducted among younger participants (e.g., Caci, Morin, & Tran, 2013; Li, Reise, Chronis-Tuscano, Mikami, & Lee, 2015; Martel, Von Eye, & Nigg, 2010) such that a two-factor bifactor model of ADHD provided the best fit to the data (Gomez et al., 2014; Gomez et al., 2013; Martel et al., 2012).

Taken together, these findings appear to provide initial support for a bifactor model of ADHD in adults, but there are some important limitations and omissions to consider before drawing any conclusions. First, it is not clear whether a two-factor or three-factor bifactor model best characterizes ADHD. That is, an important question has to do with whether hyperactivity and impulsivity should be considered distinct subdimensions of ADHD, or as a single subdimension in adults. Of the three studies that found support for a two-factor bifactor model, none of them included a three-factor bifactor model in their analyses so alternative bifactor models were not evaluated. Second, those same three studies did not use a rating scale that was developed for adults, making it unclear whether their findings can be attributed to their use of a measure that may not have been developmentally appropriate. Further, they all used the same rating scale, which could have increased the likelihood that the same pattern of results would be found. Third, not a single study has replicated this particular factor structure with a college student population. Further research sampling different periods of development, such as during the college years, will help to determine whether any age-related pattern of factor loadings replicates. Also, examining the symptom structure of ADHD for college students using a factor analytic perspective is particularly relevant at this time given the increased prevalence of ADHD in college students, rise in the number of students requesting ADHD evaluations for college students, and the recent updates made to the clinical diagnostic criteria for ADHD in adults in the DSM-5. Importantly, a bifactor model of ADHD has not been tested with the BAARS-IV. In light of these limitations, further replication is needed.
Differential Item Functioning (DIF)

One of the many advantages of IRT is that it can incorporate techniques for evaluating measurement invariance across subpopulations (i.e., whether a test item has the same relationship with theta across groups) using tests of differential item functioning (DIF; Makransky & Glas, 2013). This is an important consideration in the context of ADHD given that group differences in ADHD is another area that has not received sufficient attention in the literature. Gender differences in ADHD is one particular area that lacks adequate research (e.g., Williamson & Johnston, 2015). As in childhood, there appear to be gender differences in the prevalence of ADHD in adulthood, though national survey results show that the ratio of men to women with ADHD is less extreme than found in children, with a ratio of 1.6:1 (e.g., APA, 2013; Barkley 2006; Kessler et al., 2006). There is research to suggest that there are also gender differences in the symptom presentation of ADHD such that women are more likely than men to present primarily with inattentive features, a finding that is consistent with reports of behavioral manifestations of ADHD in children (e.g., Biederman, Faraone, Monuteaux, Bober, & Cadogen, 2004; Fedele, Lefler, Hartung, & Canu, 2012; Willcutt, 2012). For example, a recent meta-analysis found that a significantly larger proportion of females than males met criteria for the ADHD-Inattentive subtype in clinical and community samples of children (42% of females vs. 36% of males) and adults (55% of females vs. 49% of males; Willcutt, 2012), whereas males were more likely than females to meet criteria for the ADHD-Combined subtype (28% of males vs. 22% of females in children; and 26% of males vs. 18% of females in adults). Another study showed a somewhat similar pattern of results when they found that the only gender difference was in self-reported inattention, such that women reported higher symptom levels than did men (Biederman et al., 2004). A recent study that examined possible gender differences in ADHD symptoms among college students found significant differences in levels of ADHD symptoms (both inattention and hyperactivity), and impairment such that college women had higher levels of symptoms and impairment (Fedele et al., 2012). However, not all studies find gender differences in ADHD symptoms (e.g., Biederman et al., 2005; DuPaul et al., 2001; Murphy & Barkley, 1996) and consistent patterns of gender differences in ADHD symptoms have not emerged (Williamson & Johnston, 2015).

Evidence of gender differences in the presentation of ADHD symptoms raises concerns regarding whether particular items (or set of items) of the DSM-based BAARS-IV might
function differently for men and women. That is, when individuals with equal levels of the latent ADHD trait (e.g., inattention), but who differ in gender, do not have the same probability of endorsing the item, the item is considered to be exhibiting DIF. For instance, if a man and woman had the same level of inattention, but the man was more likely to endorse a specific inattention criterion, we would say that item displays DIF. The presence of DIF may indicate the existence of a problem in the validity of an item because it suggests that individuals from different subgroups of the population “have differing probabilities or likelihoods of success on an [test] item, after they have been matched on the psychological characteristic or trait of interest” (Clauser & Mazor, 1998, p. 31). Thus, when DIF is present, it implies that some of the items may not be good items for a certain group, or that different methods of scoring across those groups may be needed. Despite the widespread use of DSM-based scales for adult ADHD, published reports on this matter among adults are scarce. Smith, Jr. and Johnson (1998), using Rasch analysis, found that one DSM-IV item of inattention functioned differently across gender: “Don’t follow through on instructions and fails to finish work.” Specifically, men were significantly more likely than women to endorse this item as significant (i.e., a rating of “often” or “very often”). However, the study did not use IRT-based DIF analysis and gender differences on the discrimination parameter could not be investigated because Rasch models only include an item difficulty and the trait level for estimating item performance. Clearly, more research is needed to determine whether DIF exists between males and females (gender).

The Present Study

Overall, there is a growing body of research suggesting that the BAARS-IV is a reliable and valid tool for assessing ADHD in adults (Barkley, 2011; Becker, et al., 2014; Flannery et al., 2014), yet there are a number of limitations in existing data. The overall goal of the current study was to advance our knowledge of the measurement properties of the BAARS-IV. To this end, this study was designed to address three major objectives. The first aim was to examine the factor structure of the ADHD symptoms of the BAARS-IV. The second aim was to use an item response theory (IRT) model to examine item characteristics of the BAARS-IV in college students. The final aim was to determine whether items of the BAARS-IV were psychometrically invariant across gender.
In IRT analysis, it is necessary to first assess dimensionality. Accordingly, the first aim of the current study was to examine the dimensionality of the BAARS-IV. Although IRT-based analyses require that the dimensionality of a measure is tested before evaluation of individual item performance takes place, examining the factor structure of the BAARS-IV was important for other reasons. From a clinical and conceptual standpoint, an area of ongoing debate and research in ADHD has been the structural organization of the ADHD symptoms in terms of the number of factors, and to a lesser extent, the content of the factors. Recently, there have been studies that support the view that ADHD symptoms are best conceptualized structurally in terms of a bifactor model (Gibbins et al., 2012; Gomez et al., 2014; Gomez et al., 2013; Martel et al., 2012; Morin et al., 2013; Sorge et al., 2015). However, there have been inconsistencies across these studies in terms of whether a bifactor model with two specific factors (inattention and hyperactivity/impulsivity) or three factors (separate inattention and hyperactivity, and impulsivity) was supported. Further, there are additional limitations in previous research that preclude conclusions from being drawn.

The present study attempted to address some of these limitations by examining the factor structure of the BAARS-IV. Specifically, both correlated and bifactor models were examined to determine the best fitting model of ADHD in college students (i.e., a clinic-referred sample). To date, no study has tested a bifactor model of the BAARS-IV. Moreover, there is a lack of comprehensive data of the applicability of the bifactor model for college students. Finding that a bifactor model provides the best fit to the data could be interpreted as providing some preliminary support for the DSM-5’s focus on symptom presentation (i.e., an overall diagnosis of ADHD with inattention and hyperactivity/impulsivity as continuous modifiers) versus symptom subtypes. Further, support of such a model in young adults might allow for a more nuanced understanding of developmental changes in ADHD. Based on previous research, four models were fitted to the data, including (a) a correlated two-factor model; (b) a correlated three-factor model (with the item “talk excessively” allocated to the Impulsivity factor); (c) a bifactor model with one general ADHD factor and two specific factors (inattention and hyperactivity/impulsivity); and (d) a bifactor model of a general ADHD factor with three specific factors for residual inattention, hyperactivity, and impulsivity. The primary questions that were addressed by these analyses include:
1) Which factor model provides the best fit to ADHD symptom ratings in college students?

2) If the results of the CFA analyses support separate hyperactivity and impulsivity dimensions, a corollary question that will be tested is whether particular items (e.g., “talks excessively”) correspond more closely with the hyperactivity or impulsivity subdimension.

Based on prior work (Gibbins et al., 2012; Morin et al., 2013; Sorge et al., 2015), it was predicted that the bifactor model with a general factor and three specific factors would best account for the data.

Given that the existing psychometric information on the BAARS-IV has all been derived from the classical test theory (CTT) approach, examining the psychometric properties of the BAARS-IV using IRT-based techniques can offer new and valuable information regarding the measurement of ADHD. Additionally, less is known about the degree to which individual ADHD symptoms and the subscales (i.e., Inattention, Hyperactivity/Impulsivity) as a whole are developmentally appropriate for college students. For example, it is possible that some items are more “difficult” to endorse than others. That is, they are only present at higher levels when an individual is displaying high levels of the trait (i.e., inattention or hyperactivity/impulsivity). The second analysis of the current study consisted of applying a graded response model (GRM) to the 18 ADHD items of the BAARS-IV to acquire this type of practical data. For each item, several psychometric properties (e.g., discrimination parameters, threshold parameters, item information functions [IIFs], and category response curves) were calculated and examined to achieve this goal. In addition, the test information function (TIF) for each subscale was examined. The following research questions were addressed by this analysis:

3) How well does each item discriminate individuals among different levels of the underlying ability (i.e., item discrimination)?
   a. Which items provide the most information?

4) How reliable are the subscales or specific factors for measuring the symptoms at different levels of severity of the trait?

The final aim has to do with the assumption that the items of the scale measure the same trait(s) in members from different groups (i.e., gender). As previously summarized, there is mixed evidence to suggest that there are gender differences in the manifestations of ADHD.
Despite the widespread use of the BAARS, however, there is virtually no evidence that the items are psychometrically invariant across gender. To extend work in this area, this study conducted differential item functioning (DIF) analyses (within an IRT framework) to determine the degree to which an item (or set of items) functions similarly or differently for males and females. Further, DIF analyses can test for differences across gender in item difficulty and item discriminability. DIF on the discrimination parameter suggests that the degree to which the item discriminates an individual’s trait level depends on the group being measured (i.e., non-uniform DIF), whereas DIF on the difficulty parameter indicates that individuals with the same underlying trait level differ in the likelihood that they will endorse an item as a function of group membership (i.e., uniform DIF). Given that items may be biased in difficulty and/or discrimination, this study generated both item discrimination and difficulty parameters to identify where the DIF exists. For any significant DIF that was detected, estimates of the size of the violation were calculated. To detect DIF that is both significant and meaningful would suggest that the item(s) may not be good for both males and females or that different methods of scoring or weighting the item(s) across gender may be needed. For example, if significant and meaningful DIF were to be found for the item “talk excessively” such that it is less strongly associated with high levels of hyperactivity/impulsivity in females than in males, one possible solution for addressing this discrepancy is to weight the item differently for males and females. The investigation of measurement invariance is important given that it is an indication of the construct validity of the items, and the presence of DIF suggests that an item or set of items of the BAARS-IV have utility for only males or females in this particular population.

5) Do the items of the BAARS-IV function differently for males and females?
   a. If an item(s) shows evidence of significant DIF across gender, is the magnitude of the DIF meaningful?

Due to the exploratory nature of these analyses, no specific hypotheses were made.
CHAPTER 2

METHODS

Participants

Archival participant data consisting of 607 undergraduate students enrolled in a large, public university in the Southeastern United States were used for the purposes of the current study. All participants had previously presented at an on-campus university clinic specializing in the assessment of ADHD, learning disabilities (LD), and other academic difficulties, and then had undergone the clinic’s standard evaluation battery. Participants provided consent for their de-identified evaluation data to be utilized for research. This clinic mostly serves students who are self-referred and are seeking assessment services to address concerns regarding their academic, professional, and/or social functioning. Participants were required to be currently enrolled undergraduate students between the ages of 18 and 25, and had to have a valid data point for the gender variable.

The participant sample was 52.2% male and 47.8% female, with an overall mean age of 20.42 years ($SD = 1.83$). The majority of the sample self-identified as Caucasian (60.8%), with the remainder of the sample identifying as Hispanic (13.7%), African American (11.9%), Asian (3.0%), or other/biracial (2.6%). Forty-nine (8.1%) participants did not report their racial/ethnic background. These statistics are comparable to the racial composition of the university population from which this sample was collected (i.e., 68.1% Caucasian, 15.1% Hispanic, 9.3% African-American, 2.8% Asian, 3.3% Other, 1.4% race/ethnicity not reported; Office of Institutional Research, 2011). Class standing was as follows: 25.4% were freshman, 23.1% sophomores, 27.7% juniors, 21.6% seniors, and 2.3% did not identify their undergraduate student level. See Table 1 for sample demographic characteristics.

Mean GPA of the participants was 2.79 ($SD = .62$). The mean Brief Intellectual Ability (BIA) score from the Woodcock-Johnson Tests of Cognitive Abilities–3rd Edition (WJ III-COG; Woodcock, McGrew, & Mather, 2001b) was 100.23 ($SD = 11.76$), which falls in the average range. Based on the procedures described below, approximately 85.7% ($n = 520$) of the participants received at least one DSM-IV-TR diagnosis, with a large proportion of the sample meeting diagnostic criteria for an ADHD diagnosis (77.1%; $n = 468$). Of the clients diagnosed
with ADHD, 50.2% \((n = 235)\) were diagnosed with \textit{DSM-IV-TR} ADHD Combined Type (ADHD-C), 48.3% \((n = 226)\) were diagnosed with ADHD Predominantly Inattentive Type (ADHD-IA), and 1.1% \((n = 5)\) were diagnosed with Predominantly Hyperactive-Impulsive Type (ADHD-HI). ADHD subtype was not specified for two of the participants \(<1\%). Approximately 26% of those who received a diagnosis of ADHD also met diagnostic criteria for another disorder, with the most common additional diagnosis being an anxiety disorder \(10.9\%\).

**Measure**

**Barkley Adult ADHD Rating Scale-IV**

Current symptoms of ADHD were assessed using the Barkley Adult ADHD Rating Scale-IV: Self-Report: Current Symptoms (BAARS-IV: SR: CSS; Barkley, 2011; referred to hereafter as the BAARS-IV). The BAARS-IV is a 30-item self-report measure designed to evaluate ADHD symptoms and related functioning during the past six months in adults between the ages of 18 and 89. The BAARS-IV includes 18 items (nine items of inattention, five items of hyperactivity, and four items of impulsivity) that were developed to correspond to the diagnostic criteria for ADHD in the \textit{DSM-IV-TR}, although they were slightly modified in language to better fit adult symptoms (e.g., references to school/schoolwork are removed, “play” activities replaced with “fun”), which is more consistent the wording reflected in the \textit{DSM-5} (5th edition; APA, 2013). In addition, there are nine items evaluating sluggish cognitive tempo (e.g., easily confused, slow moving). In reference to the past six months, participants rate each of the first 27 items using a 4-point Likert scale: (1) never or rarely, (2) sometimes, (3) often, and (4) very often. The remaining three items assess additional \textit{DSM} ADHD criteria, including age of onset of endorsed symptoms and domains of impairment (i.e., home, education, work, and social). For the purposes of this study, only the 18 items that correspond to the \textit{DSM} were considered.

The BAARS-IV yields four empirically-derived subscales: Inattention (IA), Hyperactivity (HYP), Impulsivity (IMP), and sluggish cognitive tempo (SCT). Overall, the BAARS-IV Current Symptoms scale appears to have sound psychometric characteristics. Internal consistency for the BAARS-IV ADHD subscales, as reported in the manual, range from satisfactory \((\alpha = .78, \text{Hyperactivity}; \alpha = .81, \text{Impulsivity})\) to excellent \((\alpha = .90, \text{Inattention}; \alpha = .91, \text{ADHD Total Score}; \text{Barkley, 2011})\). Other studies have reported similar reliability
coefficients (Flannery et al., 2014; Voinescu et al., 2012). The BAARS-IV has also demonstrated adequate test-retest reliability over a 2- to 3-week period, with reliability coefficients ranging from .66 on Inattention to .76 on Impulsivity ($r = .66$, Inattention; $r = .72$, Hyperactivity; $r = .76$, Impulsivity; $r = .75$, ADHD Total Score). With respect to the validity of the BAARS-IV, Barkley (2011) found small to medium correlations with a continuous performance test and medium to large correlations with a self-report measure of executive functioning. The BAARS-IV has been utilized in studies involving clinical populations as well as studies of non-clinical samples (e.g. Becker et al., 2014; Flannery et al., 2014; Fleming, McMahon, Moran, Peterson, & Dreessen, 2015; Langberg, Becker, Dvorsky, & Luebbe, 2014).

Within the current investigation, the Current ADHD total score and the nine-item IA and Hyperactivity/Impulsivity (HI) subscales demonstrated good to excellent internal consistency ($\alpha$’s = .89, .85, and .84, respectively). For the group of five HYP and four IMP symptoms as specified in the BAARS-IV manual, $\alpha$’s = 74, and .82, respectively. For the group of six HYP and three IMP symptoms as specified for ADHD in the *DSM-5*, internal consistencies were also adequate ($\alpha$’s = .77 and .80, respectively).

**Procedure**

Archival clinical data were drawn from a database used to store de-identified information for college students who had previously presented at a university-based clinic and had received an evaluation assessing ADHD, LD, and related difficulties. At the time of their evaluation, participants provided written consent for their de-identified evaluation data to be used for research. Clinic evaluations were conducted by advanced graduate students who were enrolled in either a combined doctoral program in counseling psychology and school psychology or a combined M.S./Ed.S. program in school psychology, and had received coursework and training in the evaluation of adult ADHD. All evaluation cases were supervised by two doctoral-level psychologists, at least one of which was a licensed psychologist. As part of the comprehensive evaluation process, participants completed an application for services, a demographic inventory, a large battery of emotional and behavioral rating scales, a clinical interview, cognitive and achievement testing, and a structured interview. Also, participants were asked to provide a copy of their transcript (i.e., for review of GPA), previous psychological reports (when available), and to identify an informant (e.g., parents, significant others, roommates) to complete other-report
forms. Prior to the clinical interview, all participants completed the Barkley Adult ADHD Rating Scale—IV (BAARS-IV; Barkley, 2011). An examiner was present while the participant filled-out the form so that assistance could be provided as needed. Clinical diagnoses were made according to the criteria outlined in the DSM-IV-TR. Each case was reviewed by the graduate-level clinician and at least one licensed psychologist, and a consensus between the two (or three) was needed before diagnostic feedback was provided to the participant. The Institutional Review Board (IRB) at Florida State University approved the use of the aforementioned de-identified data set for this study (see Appendices A and B for the university’s IRB letters of approval).

Data Analytic Plan

Descriptive analyses and Missing Value Analysis (MVA) were conducted using IBM SPSS Statistics (version 24). A series of confirmatory factor analyses (CFAs) were conducted to test and contrast four factor structures that have been hypothesized to underlie the 18 symptoms of ADHD. The following four models were estimated: (a) a two-factor model with correlated inattention and hyperactivity/impulsivity; (b) a three-factor model with correlated inattention, hyperactivity, and impulsivity (with the item “talk excessively” allocated to the Impulsivity factor); (c) a bifactor model with one general ADHD factor and two specific factors of inattention and hyperactivity/impulsivity; and (d) a bifactor model with one general ADHD factor and three specific factors of inattention, hyperactivity, and impulsivity. Figures 1 and 2 provide visual depictions of the correlated factor and bifactor models that were tested. As noted previously, the bifactor models are distinct from the correlated models in that they posit a general factor to account for the covariation shared by all of the ADHD symptoms and separate, uncorrelated (i.e., orthogonal) specific factors representing residual covariation among the subgroups of items.

All CFA models were estimated with Mplus version 8.0 (Muthén & Muthén, 1998-2017) using the robust weighted least squares mean and variance adjusted estimator (WLSMV), as previous research supports this method to using general maximum likelihood (ML) estimation when the observed variables are categorical and data are non-normally distributed such as those used in the present study (Beauducel & Herzberg, 2006; Brown, 2015; Finney & DiStefano, 2006; Flora & Curran, 2004; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Li, 2016; Muthén, du Toit, & Spisic, 1997). Although there is preliminary evidence to support the use of
the WLSMV estimator, it has not been extensively studied and there are no agreed upon guidelines as to what signifies “good model fit” (Sass, 2011). Thus, model fit here was evaluated based on recommendations from currently available research and should be interpreted with caution.

Model fit was evaluated using multiple fit indices, including the root mean square error of approximation (RMSEA; Steiger & Lind, 1980), comparative fit index (CFI; Bentler, 1990), Tucker-Lewis Index (TLI; Tucker & Lewis, 1973), and the weighted root mean square residual (WRMR), which have been found to perform reasonably well when obtained with the WLSMV estimation technique (DiStefano & Morgan, 2014; Flora & Curran, 2004; Yu, 2002; Yu & Muthén, 2002). It has been suggested that these fit indices may be interpreted similarly to when ML estimation is used, with RMSEA values equal to lower than .06 and .08 being indicative of good and acceptable model fit, respectively, and values between .08 and .10 suggest marginal fit (Hu & Bentler, 1999; Yu, 2002). For the CFI, values between .90 and .95 were considered to be indicative of acceptable fit and values greater than .96 were considered evidence of good model fit (Yu, 2002). For the TLI, values greater than .90 and .95 were considered to be indicative of acceptable and good fit, respectively (Hu & Bentler, 1999; Yu, 2002). With respect to the WRMR, typically values less than one suggest good fit (Yu & Muthén, 2002). Although the chi-square ($\chi^2$) test is also reported, it was not used for evaluating model fit as it is sensitive to the sample size and the number of indicators per factor (Hu & Bentler, 1999; Moshagen, 2012; Mulaik, 2007). Model evaluation also included inspection of the factor loadings. Ideally, loadings are positive, significant, and salient ($\geq .40$; Brown, 2015). Guided by suggestions put forth by Chen and colleagues (2012), a salient loading for the bifactor models was defined by $\geq .40$ on the general factor and $\geq .20$ on the specific factors. To compare the difference in model fit between nested models, adjusted chi-square difference tests were computed using the DIFFTEST option for the WLSMV estimator in MPLUS (Asparouhov & Muthén, 2006; Muthén, 1998-2004; Reise, Bonifay, & Haviland, 2013). To identify the model that best represented ADHD in the current sample, the following information was considered: fit statistics, reasonable and interpretable factor loadings, and the statistical difference of the DIFFTEST.

To determine whether a 2PL model provided significantly better fit to the data than did a 1PL model, MPLUS was used to generate item parameter estimates for both models and then a chi-square comparison of log-likelihood values from each model was performed. For the 2PL
model, a graded response model (GRM; Samejima, 1969; Samejima, 1997) was applied to the data. The GRM is appropriate for use with measures that have polytomous response formats (Edelen, Thissen, Teresi, Kleinman, & Ocepek-Welikson, 2006) such as the BAARS-IV, which has ordered categorical response options (i.e., from “never or really” to “very often”). For each of the 18 items, various psychometric properties were examined. Recall that the latent construct of interest theta (θ) signifies the underlying level of inattention, hyperactivity, or impulsivity. The model yielded a single discrimination parameter or slope (a) and three difficulty parameters or thresholds (bs) for each item. Discrimination parameters estimate the degree to which an item can differentiate participants among different levels of the underlying latent trait, with larger discrimination parameters indicating that the probability of endorsing an item increases more rapidly with the increasing levels of the latent trait. For the discrimination parameter estimates, values < 0.34 were considered very low, between 0.35 and 0.64 were low, between 0.65 to 1.34 were moderate, 1.35 to 1.69 were high; and > 1.7 were very high (de Ayala, 2009; Baker, 2001). Difficulty parameters indicate the point of θ in which a participant has a 50% probability of endorsing the item above a particular item threshold. For example, the first difficulty parameter (b1) represents the amount of overall hyperactivity needed for the participant to have a 50% of endorsing the “never or rarely” response option. Negative difficulty parameters indicate that lower than average levels of the trait are required for positive endorsement in the corresponding category, whereas positive parameters reflect that higher than average levels of the trait are required.

The parameter estimates were used to estimate category response curves (CRCs), item information function (IIFs) curves, total information function (TIFs) curves, and standard error of measurement (SEM) plots. At the item level, CRCs depict the probability of endorsing a particular response category across the range of the latent trait. The discrimination parameter corresponds to the steepness of the slope of the CRC and indicates how narrowly or broadly an item can differentiate between participants with different trait levels. A high discrimination parameter value for an item indicates that the item can better discriminate between participants with different trait levels and corresponds to steeper slopes on the characteristic curves. The difficulty parameters correspond to the point at which the curve crosses the 50% probability line on the x-axis and the middle response options peak. Also at the item level, IIFs depict the amount of information provided at each point across the latent trait. That is, IIFs characterize an
item’s degree of precision in measuring different levels of the latent construct, such that higher levels of information indicate greater precision and less measurement error. At the scale level, a TIF was plotted for each scale to display the amount of information the scale provides across the latent trait range. These were estimated by summing all of the IIF values (i.e., discrimination and difficulty parameters) for all of the items within a scale. The SEM is equal to the inverse square root of the TIF and can be used as an indicator of the precision of a group of items (e.g., subscales) across different trait levels, with lower standard error denoting more information or greater precision in the subscale with respect to latent theta.

Next, DIF across gender was tested using likelihood ratio tests (LRT) as implemented in the IRTLRDIF program (version 2.0; Thissen, 2001). This particular approach can be used to detect the presence of DIF as well as identify whether the DIF is uniform (i.e., differences on the threshold parameters), non-uniform (i.e., differences on the discrimination parameters), or both. Given the multiple comparisons conducted in these analyses, Type I error was controlled for in a two-stage approach using the Linear Step-up Procedure (Benjamini & Hochberg, 1995). In addition to the statistical significance of the tests for DIF, computing measures of effect sizes for significant DIF has been shown to be useful for assessing the practical importance of DIF (Meade, 2010; Stark, Chernyshenko, & Drasgow, 2004). Moreover, estimating effect sizes for DIF was deemed important in the present study given that LRT is sensitive to larger samples (Rivas, Gabriel, Stark, & Chernyshenko, 2009), which can result in even trivial amounts of DIF being detected (Meade, 2010). Based on the parameters from the IRTLRDIF analyses, Meade’s VisualDF program (version 2.0.4; Meade, 2010) was used to inspect the relative size and impact of the detected DIF. Specifically, VisualDF was used to compute three item-level indices related to DIF on the BAARS-IV. The first index that was examined for the items was the expected score standardized difference (ESSD), which provides an estimated effect size \( (d) \). The ESSD is interpreted the same way as with Cohen’s \( d \), such that an effect size of .20 represents a small effect, .50 represents a medium effect, and .80 represents a large effect (Cohen, 1988). The two additional indices were the signed item difference in sample (SIDS) and the unsigned item difference in sample (UIDS). The SIDS represents the average unstandardized difference in the expected scores for the focal group as compared to the reference group. Lastly, the UIDS provides the absolute value of the difference in the expected scores for the focal group and does not allow for cancellation for non-uniform DIF. If the absolute values of the SIDS and UIDS are
equal, the differential functioning present is uniform across groups (i.e., one group consistently has higher expected scores than the other). However, if the UIDS is larger than the SIDS, it indicates further examination of the plots is needed.
CHAPTER 3

RESULTS

Preliminary Analyses and Descriptive Statistics

In the preliminary analyses, the Missing Value Analysis (MVA) in SPSS revealed that data were missing on 2.24% of the BAARS-IV ADHD items. Little’s (1988) Missing Completely at Random (MCAR) test was not statistically significant $\chi^2 (233), 256.20, p = .14$, indicating that the missing data were missing completely at random. Missing data were estimated using the expectation-maximization (EM) method in SPSS (Tabachnick & Fidell, 2007). Descriptive statistics were reviewed prior to conducting the CFA and IRT analyses. The range of responding was 1 to 4 for every item. Although there were no empty cells for any of the items, a fairly strong ceiling effect (i.e., with the higher two response categories of the Likert-scale accounting for approximately 90% of the cells) was observed for item 8 (“am easily distracted”). As expected, mean item scores ($M$ range = 1.86–3.53) were somewhat high for the individual items (in comparison with a community sample; Barkley, 2011). However, data screening revealed that skewness and kurtosis values fell within an acceptable range (< 2), which suggests relatively normal distributions of item responses.

Table 2 provides descriptive statistics for the BAARS-IV total score and subscale scores. As a group, participant’s inattentive, hyperactive, impulsive, and ADHD total raw scores were in the high range and met or exceeded the 93rd percentile threshold for clinical elevation as recommended by Barkley (2011). Specifically, hyperactivity ($M = 12.29, SD = 3.58$) and impulsivity ($M = 8.83, SD = 3.41$) scores fell between the 93rd and 94th percentiles, and their inattention ($M = 25.56, SD = 5.94$) and ADHD total ($M = 46.67, SD = 10.56$) scores were at the 97th percentile. An examination of the mean subscale raw scores showed statistically significant differences by gender on the BAARS-IV total score and all subscale scores. More specifically, female participants endorsed greater inattention $t(605) = 2.77, p = .006, d = 0.23$, hyperactivity $t(605) = 2.40, p = .017, d = 0.20$, impulsivity $t(605) = 2.31, p = .021, d = 0.19$, and total ADHD $t(605) = 3.12, p = .002, d = 0.26$, symptoms than did males.
**Confirmatory Factor Analyses**

Model fit statistics for the hypothesized CFA models are presented in Table 3. In the case of the correlated factor models, the CFI and TLI values indicated acceptable fit. However, the RMSEA value for the three-factor model indicated moderate fit, whereas the RMSEA value for the two-factor model suggested only a marginal fit to the data. The three-factor model showed a significantly better fit than the two-factor model ($\Delta \chi^2 = 74.91; df = 2, p < .001$). Inspection of the factor loadings for the three-factor model indicated that all items had significant, positive loadings on their respective factors, with standardized coefficients ranging from .57 to .83 (all $ps < .001$, see Figure 3). The correlations between the factors within the three-factor model were significant and ranged from .53 to .75.

Among the bifactor models, the model with two specific factors exhibited good fit in terms of the CFI and RMSEA values; however, the TLI values for the model indicated only acceptable fit. The three-factor bifactor model did not estimate properly due to a non-positive definite covariance matrix. An inspection of the results revealed that HYP item 6 (“feel restless”) contributed to the estimation problem due to a negative variance estimate. To address this issue, the procedure was repeated after restricting the residual variance of item 6 to 0 and then again after removing item 6 completely from the model; however, these adjustments did not eliminate the issue of the non-positive definite covariance matrix. Given that the estimation of the three factor bifactor model resulted in a non-admissible solution, the results are not included in Table 3. With respect to the two-factor bifactor model, completely standardized loadings on the ADHD general factor were significant ($ps < .001$) and mostly salient (range = .34 to .72). However, two IA symptoms (“don’t follow through on instructions” and “have difficulty organizing tasks”) did not have meaningful loadings ($\geq .40$) on the general factor. Over and above their associations with the ADHD general factor, all nine IA symptoms had significant ($ps < .001$) and salient ($\geq .20$) loadings on the specific IA factor (range = .35 to .72). Further, five of the nine symptoms

---

1 To test whether responses to the BAARS-IV appeared to measure a single construct, a unidimensional model of ADHD was also tested. In this model, each item only loaded on a single factor of ADHD. For statistical identification, the mean was set to zero and the factor variance to one. This model fit was poor, $\chi^2 (135) = 1587.92, p < .001$; RMSEA = .13; TLI = .76; CFI = .79. In addition, additional analyses indicated that the two-factor model showed a substantial improvement over the one-factor model ($\Delta \chi^2 = 187.16; df = 1, p < .001$). These findings suggest that a single underlying ADHD factor did not underlie the data. Thus, the results focus on presenting the fit of the correlated and bifactor models.
from the inattention symptom domain were more strongly related to the specific factor than to the general factor. In contrast, all of the symptoms of the HI factor were more strongly related to the general factor than the specific factor. In fact, only four of the HI symptoms had positive, significant ($p < .001$), and salient (range = .39 to .56) loadings on the specific HI factor. The salient HI symptoms were the impulsivity symptoms (“talk excessively,” “blurt out answers before questions have been completed,” “have difficulty awaiting turn,” “and interrupt or intrude on others”). The five symptoms that presented the lowest level of specificity (loadings ranging from $-0.26$ to $0.10$) associated with the HI specific factor were all related to hyperactive behaviors, rather than impulsivity. Three of the five loadings were negative, two of which were significant ($p < .002$). The two remaining positive loadings were neither salient nor significant on the specific HI factor.

The two-factor bifactor model demonstrated significantly better model fit than the correlated three-factor model ($\Delta \chi^2 = 143.09; df = 15, p < .001$). However, in the two-factor bifactor model, all of the loadings of the hyperactivity items on the HI specific factor were either nonsignificant or negative. The presence of negative item loadings in the two-factor bifactor model suggest that the model is not an appropriate representation of the data for this sample. An inspection of the results suggested that the hyperactivity items loaded highly on the general ADHD factor, but did not contain a sufficient amount of specific variance to form a separate HI factor. Although the two-factor bifactor model provided a statistically better fit than the correlated three-factor model, the presence of nonsignificant and negative loadings in the bifactor model indicated that at least one of the specific factors was not defined properly, which precludes meaningful interpretation of this model. The correlated three-factor model had significantly better fit than the one- and two-factor models, and all of the factor loadings were significant and within an acceptable range, suggesting this model was the most appropriate representation of the data for this sample.

**Item Response Theory Analyses**

IRT analyses were conducted using Mplus version 8.0 (Muthén & Muthén, 1998-2017). Item parameter estimates for both a 1PL model and a 2PL (i.e., GRM) model of the BAARS-IV were generated and models were then compared using a chi-square comparison of the log-likelihood values from each model. Results showed that the 2PL model provided a significantly
better model fit than a 1PL model, $\chi^2 (11) = 122.496, p < .05$ (Embretson & Reise, 2000). Thus, Samejima’s (1969) GRM was applied to the data to evaluate the psychometric properties of the three subscales of the BAARS-IV. The discrimination (also referred to as slope or $a$) and difficulty (also referred to as threshold or $b$) parameter estimates, in the logistic metric, for the IA, HYP, and IMP items are presented in Table 4. Recall that discrimination parameter estimates $< 0.34$ are considered very low, between 0.35 and 0.64 are low, between 0.65 to 1.34 are moderate, 1.35 to 1.69 are high; and $> 1.7$ are very high (de Ayala, 2009; Baker, 2001).

**Inattention Scale**

The item discrimination parameters for the IA scale ranged from 0.78 to 1.30, which suggests all of the items had a moderate degree of precision in assessing college student’s symptoms of inattention. The two items with the largest discrimination parameters were items 7 (“loses things”) and 9 (“forgetful”), indicating that these items had the greatest precision of measurement. The least discriminating items were items 3 (“don’t listen”) and 6 (“avoids tasks”). The low discrimination parameters for items 3 and 6 suggest that responses to these items were less related to inattention than the other seven items on the IA subscale. An inspection of the IIFs presented in Figure 5 (Panel A) confirmed that items 7 (“loses things”) and 9 (“forgetful”) are more discriminating across most of the of the latent trait than items 3 (“don’t listen”) and 6 (“avoids tasks”).

As shown in Table 4, the difficulty parameter estimates for the inattention items spanned ranged from $-3.02$ to $+1.78$, indicating that the items tended to be fairly well spread out over the latent trait. However, nearly all of the first and second threshold parameters were below zero, which suggests the items of this subscale were more sensitive at the lower end of the spectrum and less sensitive at moderate-to-severe levels of inattention. Furthermore, the threshold parameters for item 8 (“easily distracted”) were all negative, indicating this item was poorly spread out over the latent trait and participants with relatively low levels of inattention were still endorsing the most severe response option for this item. The difficulty parameters indicated a reasonable increase in values between response choices within each item (i.e., change in $b$ parameters ranged from 0.85 to 1.67). This can also be illustrated with the CRCs presented in Figure 6 such that the CRCs for each IA item show distinct shifts (i.e., none of the peaks for any of the response categories are overlapped by another) to higher trait levels as the response
options increase. Other than item 8 (“easily distracted”), item 1 (“make careless mistakes”) had the lowest difficulty parameter ($b_1 = -2.35$), indicating that even the respondents with below average inattention were likely to obtain an expected item score above one. Items 3 (“don’t listen”) and 7 (“lose things”) tended to be associated with the highest levels of severity (i.e., the highest $b$ parameters); for these items, higher levels of inattention were required to endorse the more severe response categories.

Figure 5 (Panel A) shows that the majority of items were most informative in the average range of the latent trait (i.e., from approximately $-1.0 \, SD$ and $+1.0 \, SD$) and were least informative at the sub-clinical range onward (i.e., $+1.5 \, SDs$ and above). Items 1 (“make careless mistakes”), 7 (“lose things”), and 9 (“forgetful”) provided the most information across a relatively broad range of the latent trait (i.e., between $-2.0 \, SDs$ and $+1.5 \, SDs$). The remaining items provided less information, though item 3 (“don’t listen”) provided relatively more information at subclinical levels of inattention ($+1.5 \, SDs$) and above. Examination of the TIF (Figure 5, Panel B), which provides an indication of the amount of measurement precision across levels of the latent trait, showed that the test information was greatest for $\theta$ values between $-1.0 \, SD$ and $+0.50 \, SD$.

**Hyperactivity Scale**

Table 4 includes the discrimination and difficulty parameters for the five HYP items. The item discrimination parameters for the HYP scale ranged from 0.79 to 1.12, indicating that these items exhibited a moderate degree of precision in discriminating participant’s severity levels across the trait continuum. The two items with the largest discrimination parameters were items 12 (“feel restless”) and 13 (“difficulty engaging in activities quietly”), indicating that these items had the greatest degree of precision in discriminating participant’s severity levels. Items 11 (“leave seat”) and 14 (“on the go”) had the lowest discrimination parameters, thereby providing the lowest degree of precision in measuring certain levels of hyperactivity. An examination of the IIFs presented in Figure 7 (Panel A) confirms that item 12 (“feel restless”) and item 13 (“difficulty engaging in activities quietly”) offered greater potential for discriminating among individuals than items 11 (“leave seat”) and 14 (“on the go”) across much of the latent trait continuum, though item 11 provided more information than the other items at high levels ($+1.5$ SDs).
SDs and above) of hyperactivity. This is also reflected with the CRCs presented in Figure 8. For instance, the slopes on the CRCs for item 14 are relatively lower, as compared to the other items.

The difficulty parameter estimates for the hyperactivity items ranged from −2.13 to 2.04 (see Table 4). However, there was wide variability across items with respect to the range of the latent trait they covered. For example, item 10 (“fidget or squirm in seat”) had the lowest difficulty parameters, with a $b_1 = -2.13$, $b_2 = -0.95$, and $b_3 = 0.10$, suggesting that this item covered θ space that was shifted toward the less severe end of the latent trait. This also means that this was the “easiest” item and students with even average levels of hyperactivity were likely to endorse the more severe response options. Item 11 (“leave seat”) had the highest difficulty parameters for $b_1 (-0.15)$, $b_2 (1.09)$, and $b_3 (2.04)$, indicating that this was the most “difficult” item and that high levels of hyperactivity were necessary for a student to choose a response option other than “1” (never or rarely). Table 4 shows there was a noticeable increase in the difficulty values between response options (change in $b$ parameters ranged from 0.77 to 1.34). This is also illustrated with the CRCs presented in Figure 8 such that the CRCs for each HYP item show distinct shifts to higher trait levels as the response options increase.

Figure 7 presents the IIFs for each of the hyperactivity items (Panel A), and the test information function (Panel B) and its standard error (Panel C) for the HYP subscale. As shown in Figure 7 (Panel A), at the item level, nearly all of the items on the HYP subscale were most informative at average levels of inattention (i.e., between −0.5 SD and +1.5 SDs) and were least informative at the upper and lower ends of the latent trait continuum. Item 10 (“fidget or squirm in seat”), however, provided the most information at the negative end of the latent trait and almost no information at the positive end of the continuum. Inspection of the TIF presented in Figure 7 (Panel B) revealed that the HYP subscale yielded the most information around mean trait levels (i.e., θ between −1.0 SD and +1.00 SD).

**Impulsivity Scale**

As shown in Table 4, the discrimination parameter estimates for the items on the IMP subscale ranged from 1.05 to 1.50. With the exception of item 15 (“talk excessively”), the items of the IMP subscale demonstrated a high degree of precision in discriminating participant’s severity levels on the underlying dimension of theta. Among the IMP items, item 15 (“talk excessively”) had the lowest discrimination parameter ($a = 1.05$), indicating this item had only a
moderate degree of precision of measurement. Inspection of the IIFs presented in Figure 9 (Panel A) also showed that item 15 ("talk excessively") offered less potential for discriminating among participants than items 16 ("blurt out answers"), 17 ("difficulty awaiting turn"), and 18 ("interrupt or intrude").

For the items on the IMP subscale, the difficulty parameters spanned from a theta value of −1.46 to 3.53. This suggests that the items on the IMP subscale were fairly well dispersed over the latent trait continuum and were most likely more sensitive in estimating above average levels of impulsivity. As shown in Table 4, there was a reasonable amount of change in the difficulty parameters between each response option within an item (change in b parameters ranged from 0.73 to 1.07). The distinct shifts to higher trait levels as the level of the response options increased can be seen in the CRCs in Figure 10. Item 15 ("talk excessively") had the lowest difficulty parameters for $b_1$ (−1.46), $b_2$ (0.41), and $b_3$ (1.80), indicating that this was the "easiest" item. Item 16 ("blurt out answers") had the highest difficulty parameters for $b_1$ (−1.08) and $b_3$ (3.53), meaning that a relatively higher level of $\theta$ was required for a participant to respond to an item with a score above 1, relative to the other items with lower difficulty values.

The IIFs for each of the impulsivity items (Panel A), and the test information function (Panel B) and its standard error (Panel C) for the IMP subscale are presented in Figure 9. As illustrated in Panel A, item 16 ("blurt out answers") provided the most information across the latent trait continuum compared to the other impulsivity items. Items 17 ("difficulty awaiting turn") and 18 ("interrupt or intrude") offered a similar amount information up to approximately $+0.5 SD$, but then item 18 begins to offer more information, especially for participants with $\theta = 1$. Item 15 ("talk excessively") provided the least amount of information. Panel B (Figure 9) shows that for the IMP scale as a whole, information tended to be relatively higher from $−0.5 SD$ to $+1.5 SD$s from the mean, compared to information below $−0.5 SD$ from the mean.

**Differential Item Functioning**

To test for the presence of DIF as a function of gender, the IRTLRDIF program (Thissen, 2001) was used. The analyses of DIF were conducted separately for each ADHD subscale (IA, HYP, and IMP subscales). The discrimination and difficulty parameters for males and females are included in Table 5. After conducting all DIF analyses, each item's item-level test was evaluated using the linear step-up procedure (Benjamini & Hochberg, 1995; Thissen, Steinberg,
& Kuang, 2002) to control for Type I error. As can be seen in Table 5, significant item-level DIF was identified for five of the 18 total items, including items 4 (“don’t follow through on instructions”) and 5 (“difficulty organizing tasks”) from the IA subscale, items 11 (“leave my seat”) and 12 (“feel restless”) from the HYP subscale, and item 15 (“talk excessively”) from the IMP subscale. Next, the specific parameter tests for these five items were evaluated. Results showed that all five items yielded significant DIF on the difficulty ($b$) parameter, indicating the DIF was uniform. With respect to the item discrimination parameters ($a$s), there were no significant differences across genders.

Next, three effect size indices for each item were calculated to examine the size and impacts of the DIF. Table 5 presents the computed effect size indices. The effect sizes (ESSDs) for items with significant DIF for gender indicated the impact of DIF on the items was moderate. Of the five items with significant DIF, the largest SIDS was for item 5 on the IA subscale (“difficulty organizing tasks”). The SIDS for item 5 was $-0.27$, indicating that the impact of DIF, when allowing for cancellation of DIF across participants, was almost 0.30. The UIDS was also 0.27, which rules out the presence of non-uniform DIF for this item because the value is equal to the value of the SIDS. The ESSD for item 5 also yielded the largest effect size ($d = -0.43$), which indicates almost one-half of a standard deviation difference and a medium effect. Given that females were selected as the focal group for these analyses, the negative values of the effect size indices for item 5 mean that for any given level of inattention, male college students scored higher on the item “difficulty organizing tasks.” A similar interpretation can also be applied for items 4 (IA scale) and 11 (HYP scale). However, for items 12 (HYP scale) and 15 (IMP scale), the pattern was reversed such that these items were “easier” for females. That is, for any given level of hyperactivity, female college students endorsed greater severity on the item “feel restless,” and for any given level of impulsivity, females scored higher on the item “excessive talking.” Inspection of the SIDS and UIDS values for items 4, 11, 12, and 15 indicated that they were similar in magnitude, which suggests DIF was consistent across the latent traits. The total score effect size (ETSSD), which represents the impact of DIF on the total scale score, indicated a small impact of DIF on the IA ($d = 0.01$), HYP ($d = 0.02$), and IMP ($d = 0.01$) subscales.
CHAPTER 4
DISCUSSION

Given the increasing number of college students seeking evaluations and services for symptoms associated with ADHD, as well as the impairment associated with the disorder as young adults, research on the measurement of ADHD in this age group is critically important. With this in mind, the present study examined the factor structure and measurement properties of a commonly used rating scale designed to assess ADHD symptoms in adults with a clinic-referred sample of college students. Specifically, the present investigation extends the previous literature on ADHD in college students by comparing recent bifactor models with more traditional correlated factor models of ADHD, using item response theory (IRT) analyses to evaluate item- and scale-level characteristics of the BAARS-IV, and testing whether these items were psychometrically invariant across gender.

Conceptualization of ADHD

One of the primary questions of this study was whether ADHD in college students is best conceptualized as having a correlated factor structure (i.e., two- or three-factor) or a bifactor structure that comprises a general factor of ADHD, as well as two (or three) uncorrelated specific factors (i.e., inattention, hyperactivity, and impulsivity). Accordingly, four potential models were evaluated using confirmatory factor analyses (CFA) to determine the best representation of the relations among inattention, hyperactivity, and impulsivity symptoms in college students. Overall, the correlated three-factor model (composed of nine items of inattention, five items of hyperactivity, and four items of impulsivity) was selected as the optimal representation of ADHD symptoms in the current sample of college students, as it showed better fit than the two-factor and three-factor (composed of nine items of inattention, six items of hyperactivity, and three items of impulsivity) models, and demonstrated more interpretable and salient factor loadings than the bifactor two-factor model. Although a similar factor structure has been reported in previous studies of adults using similar DSM-based rating scales (Barkley et al., 2008; Matte et al., 2015; Murphy & Barkley, 1996; Proctor & Prevatt, 2009), in a college student population using the BAARS-IV (Becker et al., 2014), and in a population-based study of adults using the
BAARS-IV (Barkley, 2011), the current study is one of the first to include competing bifactor models in the comparison and also extends the findings to a clinical college population. The three-factor structure closely parallels the International Classification of Diseases (ICD-10; World Health Organization, 1993) organization of ADHD such that inattention, hyperactivity, and impulsivity are each separate, albeit correlated, factors, and item 15 “talk excessively” is listed under the impulsivity dimension. Unlike factor analytic studies of childhood ADHD symptoms that have identified a two-factor model of ADHD, which is mostly consistent with the DSM-IV-TR (e.g., DuPaul et al., 1998), the findings from this study with adults suggests that the symptoms of impulsivity are primarily verbal in nature and form a semi-independent dimension of their own by young adulthood (e.g., Barkley, 2011). Although the present study did not utilize external correlates, there is substantial support for the validity of separating hyperactivity and impulsivity in adulthood (e.g., Barkley, 2011; Becker et al., 2014; Proctor & Prevatt, 2009). For example, Becker et al. (2014) examined the associations between inattention, hyperactivity, and impulsivity and sleep functioning in college students with ADHD and found that symptoms of hyperactivity (but not impulsivity) were significantly associated with increased use of sleep medication, longer sleep onset latency, poorer sleep quality, and shorter sleep duration, even after controlling for other independent variables.

In the current study, neither the two-factor bifactor model nor the three-factor bifactor model appeared to be an adequate representation of the data for this sample, as the former exhibited nonsignificant and significant negative factor loadings of the hyperactivity items onto the specific HI factor and the latter produced a non-admissible solution. The out-of-bound parameters for the two-factor bifactor model indicated this model was not valid. Closer inspection of the results indicated that the estimation problems for both models were most likely due to the hyperactivity items loading so strongly on the general factor that there was little variance left to model a specific factor. Although nearly all previous studies that have included bifactor models have supported the bifactor structure of ADHD in adults, many of them have also reported finding the presence of at least one improperly defined specific factor as evidenced by the presence of negative and/or nonsignificant loadings onto at least one of the specific factors (e.g., Gomez et al., 2014; Martel et al., 2012). The finding that HI items describing physical overactivity (i.e., hyperactivity) exhibited non-significant and negative loadings on the HI specific factor might reflect the developmental decline of these behaviors with age (e.g., Hart,
Lahey, Loeber, Applegate, & Frick, 1995; Barkley, 2006). However, considering this bifactor model fit the data significantly better than the correlated three-factor model, the finding that the HI items were more strongly related to the general construct might also indicate that the current measurement of ADHD symptoms is not adequate for capturing certain characteristics of ADHD in young adults and needs to be improved. However, due to the measurement issues for the bifactor model the interpretability of this model is limited.

**Item Response Theory**

The second aim of the current study was to evaluate the psychometric properties of the 18 ADHD symptoms found in the BAARS-IV using IRT-based methods. Few studies have examined the IRT properties of adult self-report ratings of an ADHD rating scales, and the current study, to our knowledge, is the first to examine the psychometric properties of an ADHD rating scale that is specifically used for the assessment of ADHD in adults. As such, the results of the current study may offer valuable new information regarding its items and scales. Using a graded response model, the degree of measurement precision of the BAARS-IV across participant’s inattentive, hyperactive, and impulsive behaviors was evaluated. Furthermore, whether the items in the subscales functioned similarly across gender was also investigated.

The general findings of the current study were that all of the items of the BAARS-IV had adequate (i.e., moderate or high) discrimination values, which suggests the items were adequate at discriminating different levels of the relevant latent traits (IA, HYP, or IMP). Despite this, there were some differences with respect to how well items discriminated between different levels of the traits. Overall, the items within each subscale showed a reasonable shift toward higher levels of the relevant latent trait as response options increased, indicating that more of the relevant trait was required to endorse higher ratings. The items and their respective traits tended to be most informative for participants with trait levels between $-0.50 \, SD$ and $+0.50 \, SD$, but they were less reliable at identifying individuals at high and very high levels of the latent trait.

The results of the GRM analyses for the IA subscale revealed that all of the items had moderate discrimination values, suggesting adequate fit to the latent trait. However, there was some variability with respect to how well the items were at discerning different levels of inattention. For example, items 3 (“don’t listen”) and 6 (“avoids tasks”) were found to have relatively low discrimination values, indicating these items had less discriminating power than
the other IA items. Items 7 (“loses things”) and 9 (“forgetful”) had the highest discrimination values and were therefore the best items for differentiating certain levels of inattention in college students. The difficulty parameter estimates for the IA items spanned −3.02 to +1.78, which suggests that the items were adequately dispersed over the latent trait continuum. However, nearly all of the first and second threshold parameters for the items were below zero indicating that at even below average and average levels of inattention, participants were likely to endorse these items at a level above “never or rarely.” The threshold parameters suggest that for the IA items, item 3 (“don’t listen”) had the highest thresholds, whereas item 8 (“easily distracted”) had the lowest threshold values. This latter finding is consistent with studies with children that have also reported that item 8 (“easily distracted”) was endorsed at relatively lower levels of IA (Gomez, 2008; Gomez, 2011; Purpura, Wilson, & Lonigan, 2010). Based on an inspection of the IIFs presented in Figure 5, nearly all of the items provided adequate information at the lower end and middle of the IA trait continuum and relatively less information at the higher end. The TIF for the IA subscale demonstrated a similar pattern such that the greatest amount of information was provided around average levels of theta (around −0.50 SD). Further, there was relatively better reliability (.88 reliability) for the subscale at latent trait levels between −1 SD below the mean trait level to approximately +.75 SD above the mean trait level. This subscale had a reliability of at least .80 from −2.5 SDs to +1.5 SD. In other words, the IA subscale has adequate reliability in the low to slightly above average range of inattention, but does not appear to reliably assess high levels of inattention.

Similar to the findings for the IA subscale, there was some variability across the discrimination and difficulty parameters across the items within the HYP scale. All of the items on the HYP subscale demonstrated adequate discrimination (i.e., had values that fell within the moderate range). However, items 11 (“leave seat”) and 14 (“on the go”) had relatively lower discrimination values than the other items, thereby indicating that they were somewhat less effective at discriminating different levels of hyperactivity. Items 12 (“feel restless”) and 13 (“difficulty engaging in activities quietly”) had discrimination parameters > 1.0, suggesting they were generally good at discriminating different levels of the hyperactivity trait. Results also suggested that the five-item HYP subscale covered a wide range of hyperactivity severity (i.e., b parameters ranged from −2.13 to 2.04). However, there was wide variability across items with respect to the range of hyperactivity severity they covered. Items 10 (“fidget or squirm in seat”)
and 12 (“feel restless”) had the lowest difficulty parameters, indicating that the item was not well represented at higher levels of the trait and it took only low levels of overall hyperactivity to endorse a response option greater than “1” (never or rarely). Item 11 (“leave seat”) was the most severe symptom amongst the HYP items. This particular finding could reflect a potential developmental decline of the importance of this symptom during adulthood. However, higher threshold values for item 11 (“leave seat”) have been reported in both child and adult samples, which would suggest this item is associated with higher severity, regardless of age (Gomez, 2008; Gomez, 2011). At the item level, items 12 (“feel restless”) and 13 (“difficulty engaging in activities quietly”) provided the most information between −1.0 SD and +2.5 SDs of the latent trait. The TIF indicated that information for the HYP subscale as a whole peaked around the mean of the latent trait; however, the subscale had a reliability level below .80 across all levels of the latent trait. This suggests that the HYP subscale does not reliably identify participants at low, average, or high levels of the latent trait.

The IMP subscale provided the greatest discriminatory capacity among the three ADHD subscales, as indicated by all discrimination parameters being > 1.0. With the exception of item 15 (“talk excessively”), all of the items had high discrimination values, thereby indicating that they were good at discriminating different levels of the latent trait. Among the four IMP items, item 15 (“talk excessively”) had the lowest discrimination parameter, meaning that it has worse discrimination ability than the other items. The best item for discriminating the levels of the latent trait was item 16 (“blurt out answers”). The difficulty parameter estimates for the IMP items ranged from −1.46 to +3.56, which indicates that the items covered an adequate range of the latent trait continuum. Item 15 (“talk excessively”) had the lowest difficulty parameters and was not as well spread out over the latent trait as the other items. Other studies have also reported similar findings such that “talk excessively” is less informative compared with the other items (Gomez et al., 2008; Li et al., 2015; Purpura et al., 2010). This is particularly noteworthy because these studies were based on non-clinical and clinical samples of children and different measures were used. The item that tended to have the highest difficulty parameters was item 16 (“blurt out answers”), indicating that participants had to possess higher levels of impulsivity to be more likely to endorse response option greater than “1” (never or rarely). The IIFs demonstrated that, with the exception of item 15 (“talk excessively”), the information the items provided tended to be most reliable between −0.50 SD to +1.5 SDs from the mean. At the
subscale level, information peaked around +0.50 SD from the mean and the subscale had a reliability level of at least .80 from −1.0 SD to +1.5 SDs from the mean. This means that the IMP subscale has adequate reliability in the average to somewhat above average range of impulsivity.

**ADHD Symptom Differences by Gender**

Previous studies have shown evidence of gender difference in the presentation of ADHD symptoms (e.g., Biederman et al., 2004; Fedele et al., 2012; Willcutt, 2012); however, few studies have utilized IRT-based approaches to test the degree to which ADHD items function similarly for males and females. The present study used likelihood ratio tests to test for DIF across gender in adults. The current evaluation of measurement invariance in IRT item parameters revealed that after adjusting for Type I error, there was significant item bias by gender for five of the 18 ADHD items, including items 4 (“don’t follow through on instructions”) and 5 (“difficulty organizing tasks”) from the IA subscale, items 11 (“leave my seat”) and 12 (“feel restless”) from the HYP subscale, and item 15 (“talk excessively”) from the IMP subscale. Further, all five of these items evidenced a moderate amount of DIF, suggesting potentially meaningful differences. The significant DIF was present only on the difficulty (b) parameters, indicating the DIF was uniform. That is, among a clinic-referred sample, males and females at the same level of the latent trait (i.e., inattention, hyperactivity, or impulsivity) did not have the same probability of endorsing the same response option (i.e., the relative difficulty of certain items varied by gender). Although the DIF was uniform across all three subscales on the BAARS-IV, the direction of the item-level bias was not consistent such that it was not always in favor of one gender or the other. Specifically, three of the items identified to have significant DIF were “easier” for male students than they were for female students (i.e., items 4 and 5 on the IA subscale and item 11 on the HYP subscale) and two of the items with DIF were “easier” for female students than they were for male students (i.e., item 12 on the HYP subscale and item 15 on the IMP subscale), when the students were equated on level of the latent trait. Thus, item-level group differences found for specific HYP items might balance one another out at the subscale level, but the same cannot be said for the DIF detected for the IA and IMP items.

The finding that the IA item 4 (“don’t follow through on instructions”) was more likely to be endorsed by males relative to females, even when their overall mean level of inattention symptomatology on the latent factor was held constant, fits with other research that investigated
gender differences on a similar DSM-based ADHD rating scale in adults, though the authors did not utilize DIF within an IRT framework (Smith, Jr. & Johnson, 1998). Notably, previous studies with children also found similar results regarding the presence of DIF by gender. Specifically, Makransky and Bilenberg (2014) found that parents of boys were more likely to endorse item 11 (“leave his/her seat”) and parents and teachers of girls were more likely to endorse item 15 (“talk excessively), despite participants having the same level of the relevant latent trait. Given that items 11 and 15 have been found to display DIF by gender in both child and adult samples suggests that the item bias is evident across the developmental stages and these items may need to be weighted differently for males and females, regardless of age.

**Clinical Implications**

Although this study was not a direct investigation of diagnostic subtypes (DSM-IV-TR) or presentations (DSM-5), the findings do provide information that is relevant to question of whether or not the items comprising the “Hyperactive/Impulsive” presentation should be combined. The findings from this study support the notion that three factors, rather than two factors or a general factor, underlie the 18 DSM-based items of the BAARS-IV and that the impulsivity symptoms should not be combined with hyperactivity symptoms. That is, hyperactivity and impulsivity appear to be two separate constructs, and thus, should be measured separately. This finding provides some support for the DSM-5 in that discrete modifiers rather than a dimensional spectra are used; however, the current DSM-5 conceptualization of ADHD is different in that there is no “hyperactivity” modifier. Furthermore, based on the findings of the current study, the symptom “talks excessively” may be a better indicator of impulsivity rather than hyperactivity. As it currently stands, the DSM-5 based conceptualization of ADHD may be inappropriate for college students. Consequently, current DSM-5 ADHD presentation modifiers may be more misleading than helpful, as the different presentation types may have different treatment implications. Span, Earleywine, and Strybel (2002) posited that four separate subtypes/presentations (i.e., Inattentive, Hyperactive, Impulsive, and Combined) may be a better representation of ADHD in adults. This would suggest that a client could present with clinically significant levels of impulsivity only, without significant levels of inattention or hyperactivity, and receive a diagnosis of ADHD. However, considering the limited number of impulsivity symptoms and lack of evidence regarding the validity of four subtypes, the addition of a fourth
subtype seems unlikely. It would be beneficial for future studies to replicate and extend the findings from the current study by incorporating the informant and childhood symptoms versions of the BAARS-IV. It warrants caution that CFA is only one step towards investigating the appropriateness of applying the current DSM-5 presentation types to college students. Future research should be conducted to determine whether the findings from the current study generalize to all young adults, including those who are not in college.

The findings from the current study have potentially important implications for the use of the BAARS-IV with college students. In a clinical setting, the BAARS-IV is commonly used to screen individuals to help determine whether comprehensive clinical assessment is warranted or as part of a comprehensive assessment in diagnosing. As such, desirable items for this self-report measure are those that have high reliability at higher levels of the latent ADHD trait than is typical among similarly-aged individuals. Difficulty levels of 1.5 SDs or 2 SDs above the mean have been used to indicate the lower bound for clinically elevated behavior, as they correspond to common cut-offs used to determine clinical significance (de Ayala, 2009; Barkley, 2011). Items that provide information only at the lower ends of the inattention, hyperactivity, or impulsivity traits (i.e., normative behavior) could possibly promote over-identification of students due to their low degree of precision in measuring high levels of the latent trait (Wakschlag et al., 2007). Results of this study revealed that only one IA item (“don’t listen), two HYP items (“leave my seat” and “difficulty engaging in activities quietly”), and the four IMP items had an upper threshold that exceeded +1.5 SDs, which indicates that less than half of the items of the BAARS-IV measure higher levels of ADHD behaviors. Of the seven items that measured sub-clinical to clinical levels of the latent traits, the IA item (“don’t listen) and one of the HYP items (“leave my seat”) provided relatively less information (as < 1), suggesting less than optimal precision in the sub-clinical to clinical range. There are several reasons why items that provide low information may be weak indicators of ADHD-related behaviors in college students, including low relevance of the described behavior for this age group, differences in the way the item is interpreted among college students, and lack of measurement invariance. The remaining five items offered good measurement information around sub-clinical to clinical levels of their respective trait (i.e., hyperactivity or impulsivity). These items, in particular, may offer more relevant information than the others regarding clinical severity. In contrast, the low level of information for the other items, particularly the HYP items, suggests that they may not reliably
identify college students at high levels of their respective trait (i.e., hyperactivity or inattention). Thus, clinicians are strongly encouraged to supplement these ratings by including multiple methods and respondents to corroborate these ratings. Additional components may include, but are not limited to, clinical interviews with the student and significant others (e.g., parents, roommates, significant others), other-report ratings scales, neuropsychological testing, and a review of the client’s previous diagnostic evaluations.

Based on the findings indicating that there is variability in the measurement precision across items, clinicians may want to consider giving more credence to certain items than others. Li et al. (2015) suggested the development of a diagnostic algorithm that would consist of weighting each ADHD item according to their ability to discriminate individuals at high levels of the latent trait (i.e., items with high discrimination and information values are more heavily weighted). Such an approach has the potential to increase the precision of assessment measures, but would likely require further investigation with large samples using IRT and may not appropriate in all cases of poor measurement. The low informational capacity found for most of the IA and HYP items might also indicate the need for these items to be revised (e.g., rewording) or replaced with new items that more accurately capture ADHD in adults. Although this study does not suggest how these items could be reworded to improve their informational capacity and reliability, future research may benefit from further examining the psychometric properties of the BAARS-IV items to help determine whether revision or removal of particular items from the scale is necessary.

Items that offer consistent and precise measurement of ADHD constructs across certain subgroups of individuals are essential for appropriate identification and treatment of ADHD in college students. The present study does not support measurement invariance across the entirety of the BAARS-IV, as findings demonstrated uniform DIF across gender for one-third of the items and the magnitude of the effect size was moderate. Although additional studies are needed to replicate this finding, it offers preliminary support for the notion that clinicians may need to be cautious in interpreting certain ADHD symptoms the same for both genders. That is, because the probability of item endorsement differs as a function of gender, the BAARS-IV may provide inaccurate estimates of ADHD severity for males and females, which could impact clinical findings. This is particularly relevant for items 4 (“don’t follow through on instructions”), 5
(“difficulty organizing tasks”), and 15 (“talk excessively”), but also possibly for two of the HYP items (i.e., “leave my seat” and “feel restless”).

One possible solution for addressing these discrepancies is to weight these items differently for males and females. For example, because item 4 (“don’t follow through on instructions”) is more strongly associated with high levels of inattention in males than in females, clinicians may wish to count this item more seriously if it is endorsed by a female. As mentioned previously, studies involving children (Makransky & Bilenberg, 2014) and adults (Smith, Jr. & Johnson, 2000) have also found evidence of DIF for some of the same items reported here (items 4, 11, and 15, in particular), which offers some additional support for weighting these items differently for males and females. These similarities in findings also suggest that gender equivalency may not be fully supported for the ADHD symptoms included in other DSM-based ratings scales. Thus, a future target for future research should include determining the appropriate item weights across different groups. A more simple approach to weight items differently would be to utilize gender-specific norms and cutoff scores to inform clinical decisions when using the BAARS-IV. However, gender-specific normative data are not currently available for the BAARS-IV, which limits the usefulness of the BAARS-IV in clinical practice, particularly when working with college students. For this reason, clinicians may prefer to administer an adult ADHD rating scale that provides score profiles that are specific to gender (e.g., the CAARS). Considering the overlap of information obtained via DSM-based ADHD ratings scales and clinical interviews (e.g., Toplak et al., 2009), the current study’s findings related to DIF suggest that future revisions of the DSM diagnostic criteria for ADHD will need to consider different manifestations of the disorder by gender.

**Limitations and Future Directions**

The findings and conclusions in this study should be considered with some limitations in mind. First, the purpose of this study was to examine the psychometric properties of a specific self-report measure: the BAARS-IV Current Symptoms Scale. Although the BAARS-IV shares many similarities with some of the other DSM-based adult ADHD rating scales, they are not exactly the same (e.g., differences in the wording of items), and thus, the results of this study cannot be assumed to generalize to other DSM-based adult ADHD rating scales. In light of the valuable psychometric information that IRT can provide (Embretson & Reise, 2000), it will be
useful for future studies to investigate the psychometric properties of other versions of the BAARS-IV (e.g., self-report of childhood symptoms, informant ratings of current/childhood symptoms) as well as other adult ADHD ratings scales. Second, this study was based on a secondary data analysis, and thus, additional variables that could have been used to examine the convergent and discriminant validity of the BAARS-IV were not available. However, it is worth noting that past studies have found consistent patterns of relations for IA, HYP, and IMP factors with external constructs, providing some support for the distinction between these symptom domains (e.g., Barkley, 2011). In addition, the data utilized for this study were cross-sectional, which precluded the examination of the stability of the three-factor latent structure or IRT properties at different time points. Given the evidence of the developmental heterogeneity of ADHD (e.g., Barkley, 2006; Barkley, 2011; Willcutt et al., 2012), it is possible that there are changes to the latent structure and the amount of information the items of the BAARS-IV provide at different stages of adulthood (e.g., young adulthood, adulthood, middle age). Longitudinal IRT studies that include the BAARS-IV or other DSM-based measures would allow researchers to examine any changes over time.

A strength of this study is the examination of whether the measurement properties hold across gender; however, the DIF analyses was limited such that it did not include other sociodemographic variables of interest. This was mostly due to sample sizes that were small and unequal across certain groups (Embretson & Reise, 2000), as well as incomplete demographic data. Future research may want to consider, for example, testing whether and how the BAARS-IV items, as well as items from similar DHD ratings scales used to assess ADHD in adults, might behave differently across racial/ethnic, age groups, and raters. Items that demonstrate stability and a high degree of measurement precision across these various groups could be emphasized in future revisions of adult ADHD ratings scales and the DSM. Finally, the CFA analyses were based on the WLSMV estimator, which has been shown to perform well with ordered categorical data such as those used in this study (e.g., Flora & Curran, 2004). However, the findings and conclusions should be interpreted with some caution as the CFA analyses were somewhat limited by the available research on the WLSMV estimator, as there are no generally accepted standards for model fit when WLSMV is used for categorical data. The WLSMV estimator is commonly used in similar studies, and thus, comparisons can still be made (e.g., Gomez, 2011; Morin, 2013). Despite the aforementioned limitations, the present study contributes to the knowledge
base regarding the measurement utility of the BAARS-IV, particularly with college students. Further, the findings from this study may provide insight in the conceptualization of ADHD in young adulthood, which represents a critical developmental period between adolescence and adulthood.

**Conclusions**

Overall, the current study provides support for a correlated three-factor model of ADHD (with four impulsivity symptoms) in college students, as compared to a correlated two-factor model and bifactor models of ADHD. This study provides an important extension and replication of the literature on the latent structure of the BAARS-IV by providing some additional support for the distinction between hyperactivity and impulsivity in a large clinic-referred sample of college students (Barkley, 2011; Becker et al., 2014). It is important to note that the two-factor bifactor model fit the data significantly better than the correlated three-factor model, which is consistent with other recent studies of adult ADHD (Gomez et al., 2013; Martel et al., 2012; Matte et al., 2015); however, in the two-factor bifactor model, the hyperactivity items had either negative or negligible loadings on the HI specific factor. These findings call into question the validity and utility of bifactor models of ADHD for this particular population. Given the mixed findings across studies, it will be important for future studies attempting to replicate the factor structure of the BAARS-IV or another adult ADHD ratings scales to include correlated-factor models as well as recent bifactor models of ADHD. The preliminary psychometric evidence suggested that the BAARS-IV may be a useful instrument to screen for ADHD symptomatology in college students (Barkley, 2011). Results of the GRM analyses suggested that nearly all of the inattention and hyperactivity items demonstrated sufficient reliability at measuring students in the low to average range of their relevant latent trait, but not the sub-clinical to clinical range. All four impulsivity items and one hyperactivity item had good discrimination and difficulty parameters, providing useful information at sub-clinical levels of the latent traits. Thus, these items appear to have the most clinical utility in the assessment of ADHD in college students. This study also revealed gender differences with respect to five of the BAARS-IV items (i.e., two inattention items, two hyperactivity items, and one impulsivity item). Although DIF appeared to have had little impact on the overall subscale scores, the effect at the item level was moderate, and thus, these items should be interpreted with caution as they do not function the same for both genders. This finding suggests that if future research studies show similar results
in terms of statistical and meaningful differential item functioning, then revisions of ADHD rating scales, and potentially the *DSM*-criteria themselves, may be needed to address gender bias in the manifestation of certain indicators of the disorder. Taken together, the results of the current study suggest that, in general, self-report ratings of inattention and hyperactivity should be interpreted with caution, as many of the items do not reliably identify individuals with high levels of the latent trait. This study also highlights the need for future research on the validity and appropriate use of *DSM*-based ADHD ratings with young adults.
<table>
<thead>
<tr>
<th>Demographic Characteristics</th>
<th>n</th>
<th>Valid N %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>87</td>
<td>14.3</td>
</tr>
<tr>
<td>19</td>
<td>137</td>
<td>22.6</td>
</tr>
<tr>
<td>20</td>
<td>120</td>
<td>19.8</td>
</tr>
<tr>
<td>21</td>
<td>116</td>
<td>19.1</td>
</tr>
<tr>
<td>22</td>
<td>57</td>
<td>9.4</td>
</tr>
<tr>
<td>23</td>
<td>40</td>
<td>6.6</td>
</tr>
<tr>
<td>24</td>
<td>34</td>
<td>5.6</td>
</tr>
<tr>
<td>25</td>
<td>16</td>
<td>2.6</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>317</td>
<td>52.2</td>
</tr>
<tr>
<td>Female</td>
<td>290</td>
<td>47.8</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>369</td>
<td>66.1</td>
</tr>
<tr>
<td>African American</td>
<td>72</td>
<td>12.9</td>
</tr>
<tr>
<td>Asian</td>
<td>18</td>
<td>3.2</td>
</tr>
<tr>
<td>Hispanic</td>
<td>83</td>
<td>14.9</td>
</tr>
<tr>
<td>Other</td>
<td>16</td>
<td>2.9</td>
</tr>
<tr>
<td><strong>Year in school</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>154</td>
<td>26.0</td>
</tr>
<tr>
<td>Sophomore</td>
<td>140</td>
<td>23.6</td>
</tr>
<tr>
<td>Junior</td>
<td>168</td>
<td>28.3</td>
</tr>
<tr>
<td>Senior</td>
<td>131</td>
<td>22.1</td>
</tr>
</tbody>
</table>
Table 2

Means and Standard Deviations for the BAARS-IV Subscales Raw Scores for the Total Sample and Gender Separately

<table>
<thead>
<tr>
<th>Subscales</th>
<th>Overall Sample Mean (SD)</th>
<th>Males (n = 317) Mean (SD)</th>
<th>Females (n = 290) Mean (SD)</th>
<th>Possible range</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD: Total</td>
<td>46.67 (10.56)</td>
<td>45.40 (11.0)</td>
<td>48.06 (9.90)</td>
<td>18–72</td>
</tr>
<tr>
<td>Inattention (9 items)</td>
<td>25.56 (5.94)</td>
<td>24.93 (6.16)</td>
<td>26.26 (5.62)</td>
<td>9–36</td>
</tr>
<tr>
<td>Hyperactivity (5 items)</td>
<td>12.29 (3.58)</td>
<td>11.95 (3.67)</td>
<td>12.65 (3.44)</td>
<td>5–20</td>
</tr>
<tr>
<td>Impulsivity (4 items)</td>
<td>8.83 (3.41)</td>
<td>8.52 (3.33)</td>
<td>9.16 (3.47)</td>
<td>4–16</td>
</tr>
<tr>
<td>HYP + IMP (9 items)</td>
<td>21.12 (6.20)</td>
<td>20.47 (6.39)</td>
<td>21.81 (5.91)</td>
<td>9–36</td>
</tr>
</tbody>
</table>

Note. Data are raw scores. ADHD = attention-deficit/hyperactivity disorder; HYP = hyperactivity; IMP = impulsivity.
Table 3

**Model Fit Statistics for Alternative Models of ADHD Symptoms**

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>WRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlated two-factor</td>
<td>743.25*</td>
<td>134</td>
<td>.91</td>
<td>.90</td>
<td>.09</td>
<td>1.71</td>
</tr>
<tr>
<td><strong>Correlated three-factor</strong></td>
<td><strong>587.27</strong>*</td>
<td><strong>132</strong></td>
<td><strong>.93</strong></td>
<td><strong>.92</strong></td>
<td><strong>.08</strong></td>
<td><strong>1.48</strong></td>
</tr>
<tr>
<td>Two-factor bifactor</td>
<td>409.12*</td>
<td>117</td>
<td>.96</td>
<td>.94</td>
<td>.06</td>
<td>1.10</td>
</tr>
<tr>
<td>Three-factor bifactor$^a$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note.* Parameters were estimated using the weighted least squares mean and variance (WLSMV) estimator. The overall best fitting model is indicated in bold. $\chi^2$ = chi-square test of model fit; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; WRMR = weighted root mean square residual.

$^a$ Non-admissible solution.

*$p < .001.$
Table 4

Item Parameters for the BAARS-IV in the Logistic Metric

<table>
<thead>
<tr>
<th>Item</th>
<th>Content</th>
<th>Trait</th>
<th>Discrimination $a$</th>
<th>Difficulty 1 $b_1$</th>
<th>Difficulty 2 $b_2$</th>
<th>Difficulty 3 $b_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Make careless mistakes</td>
<td>IA</td>
<td>1.03</td>
<td>-2.35</td>
<td>-0.68</td>
<td>0.49</td>
</tr>
<tr>
<td>2</td>
<td>Difficulty sustaining attention</td>
<td>IA</td>
<td>0.91</td>
<td>-1.94</td>
<td>-0.71</td>
<td>0.64</td>
</tr>
<tr>
<td>3</td>
<td>Don’t listen when spoken to directly</td>
<td>IA</td>
<td>0.85</td>
<td>-1.07</td>
<td>0.43</td>
<td>1.78</td>
</tr>
<tr>
<td>4</td>
<td>Don’t follow through on instructions</td>
<td>IA</td>
<td>0.86</td>
<td>-1.69</td>
<td>-0.19</td>
<td>1.05</td>
</tr>
<tr>
<td>5</td>
<td>Difficulty organizing tasks and activities</td>
<td>IA</td>
<td>0.92</td>
<td>-1.60</td>
<td>-0.41</td>
<td>0.67</td>
</tr>
<tr>
<td>6</td>
<td>Avoid tasks</td>
<td>IA</td>
<td>0.78</td>
<td>-1.76</td>
<td>-0.46</td>
<td>0.67</td>
</tr>
<tr>
<td>7</td>
<td>Lose things</td>
<td>IA</td>
<td>1.22</td>
<td>-1.17</td>
<td>-0.10</td>
<td>0.75</td>
</tr>
<tr>
<td>8</td>
<td>Easily distracted</td>
<td>IA</td>
<td>0.92</td>
<td>-3.02</td>
<td>-1.87</td>
<td>-0.56</td>
</tr>
<tr>
<td>9</td>
<td>Forgetful</td>
<td>IA</td>
<td>1.30</td>
<td>-1.71</td>
<td>-0.51</td>
<td>0.44</td>
</tr>
<tr>
<td>10</td>
<td>Fidget or squirm in seat</td>
<td>HYP</td>
<td>0.90</td>
<td>-2.13</td>
<td>-0.95</td>
<td>0.10</td>
</tr>
<tr>
<td>11</td>
<td>Leave my seat</td>
<td>HYP</td>
<td>0.87</td>
<td>-0.15</td>
<td>1.09</td>
<td>2.04</td>
</tr>
<tr>
<td>12</td>
<td>Feel restless</td>
<td>HYP</td>
<td>1.04</td>
<td>-1.58</td>
<td>-0.29</td>
<td>0.81</td>
</tr>
<tr>
<td>13</td>
<td>Difficulty engaging in activities quietly</td>
<td>HYP</td>
<td>1.12</td>
<td>-0.42</td>
<td>0.74</td>
<td>1.50</td>
</tr>
<tr>
<td>14</td>
<td>On the go</td>
<td>HYP</td>
<td>0.79</td>
<td>-1.15</td>
<td>0.19</td>
<td>0.97</td>
</tr>
<tr>
<td>15</td>
<td>Talk excessively</td>
<td>IMP</td>
<td>1.05</td>
<td>-1.46</td>
<td>0.41</td>
<td>1.80</td>
</tr>
<tr>
<td>16</td>
<td>Blurt out answers</td>
<td>IMP</td>
<td>1.50</td>
<td>-1.08</td>
<td>1.26</td>
<td>3.53</td>
</tr>
<tr>
<td>17</td>
<td>Difficulty awaiting turn</td>
<td>IMP</td>
<td>1.40</td>
<td>-1.34</td>
<td>0.84</td>
<td>3.05</td>
</tr>
<tr>
<td>18</td>
<td>Interrupt or intrude</td>
<td>IMP</td>
<td>1.46</td>
<td>-1.18</td>
<td>1.53</td>
<td>3.37</td>
</tr>
</tbody>
</table>

Note. IA = inattention; HYP = hyperactivity; IA = impulsivity. $N = 607$. 
Table 5

Discrimination Parameters and Threshold Values Based on Gender, Significance Levels, and Effect Sizes of DIF

<table>
<thead>
<tr>
<th>Item</th>
<th>Trait</th>
<th>Males</th>
<th></th>
<th></th>
<th></th>
<th>Females</th>
<th></th>
<th></th>
<th></th>
<th>Effect Size Indices</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$a$</td>
<td>$b_1$</td>
<td>$b_2$</td>
<td>$b_3$</td>
<td>$a$</td>
<td>$b_1$</td>
<td>$b_2$</td>
<td>$b_3$</td>
<td>$p$</td>
<td>ESSD</td>
</tr>
<tr>
<td>1</td>
<td>IA</td>
<td>1.82</td>
<td>-2.08</td>
<td>-0.48</td>
<td>0.55</td>
<td>1.83</td>
<td>-2.19</td>
<td>-0.63</td>
<td>0.61</td>
<td>0.59</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>IA</td>
<td>1.65</td>
<td>-1.54</td>
<td>-0.45</td>
<td>0.76</td>
<td>1.46</td>
<td>-2.23</td>
<td>-0.78</td>
<td>0.71</td>
<td>0.08</td>
<td>0.26</td>
</tr>
<tr>
<td>3</td>
<td>IA</td>
<td>1.49</td>
<td>-0.97</td>
<td>0.66</td>
<td>2.03</td>
<td>1.36</td>
<td>-0.97</td>
<td>0.39</td>
<td>1.76</td>
<td>0.10</td>
<td>0.22</td>
</tr>
<tr>
<td>4</td>
<td>IA</td>
<td>1.57</td>
<td>-1.81</td>
<td>-0.19</td>
<td>1.11</td>
<td>1.76</td>
<td>-1.10</td>
<td>0.06</td>
<td>1.05</td>
<td>&lt;0.01*</td>
<td>-0.26</td>
</tr>
<tr>
<td>5</td>
<td>IA</td>
<td>1.75</td>
<td>-1.69</td>
<td>0.46</td>
<td>0.61</td>
<td>1.84</td>
<td>-1.01</td>
<td>-0.07</td>
<td>0.84</td>
<td>&lt;0.01*</td>
<td>-0.43</td>
</tr>
<tr>
<td>6</td>
<td>IA</td>
<td>1.45</td>
<td>-1.62</td>
<td>0.25</td>
<td>0.84</td>
<td>1.26</td>
<td>-1.58</td>
<td>-0.45</td>
<td>0.68</td>
<td>0.20</td>
<td>0.12</td>
</tr>
<tr>
<td>7</td>
<td>IA</td>
<td>2.38</td>
<td>-1.06</td>
<td>0.03</td>
<td>0.81</td>
<td>2.21</td>
<td>-0.90</td>
<td>-0.01</td>
<td>0.83</td>
<td>0.41</td>
<td>-0.04</td>
</tr>
<tr>
<td>8</td>
<td>IA</td>
<td>1.79</td>
<td>-2.43</td>
<td>-1.46</td>
<td>-0.32</td>
<td>1.22</td>
<td>-4.36</td>
<td>-2.37</td>
<td>-0.68</td>
<td>0.03</td>
<td>0.30</td>
</tr>
<tr>
<td>9</td>
<td>IA</td>
<td>2.64</td>
<td>-1.49</td>
<td>0.34</td>
<td>0.52</td>
<td>2.28</td>
<td>-1.50</td>
<td>-0.44</td>
<td>0.52</td>
<td>0.70</td>
<td>0.03</td>
</tr>
<tr>
<td>10</td>
<td>HYP</td>
<td>1.70</td>
<td>-1.74</td>
<td>-0.73</td>
<td>0.11</td>
<td>1.58</td>
<td>-2.16</td>
<td>-0.88</td>
<td>0.31</td>
<td>0.12</td>
<td>-0.02</td>
</tr>
<tr>
<td>11</td>
<td>HYP</td>
<td>1.76</td>
<td>-0.13</td>
<td>0.93</td>
<td>1.98</td>
<td>1.08</td>
<td>0.07</td>
<td>1.70</td>
<td>2.61</td>
<td>&lt;0.01*</td>
<td>-0.31</td>
</tr>
<tr>
<td>12</td>
<td>HYP</td>
<td>2.12</td>
<td>-1.29</td>
<td>-0.04</td>
<td>1.05</td>
<td>2.59</td>
<td>-1.23</td>
<td>-0.26</td>
<td>0.58</td>
<td>&lt;0.01*</td>
<td>0.33</td>
</tr>
<tr>
<td>13</td>
<td>HYP</td>
<td>1.71</td>
<td>-0.33</td>
<td>0.78</td>
<td>1.70</td>
<td>1.61</td>
<td>-0.35</td>
<td>0.99</td>
<td>1.73</td>
<td>0.46</td>
<td>-0.08</td>
</tr>
<tr>
<td>14</td>
<td>HYP</td>
<td>1.43</td>
<td>-1.12</td>
<td>0.24</td>
<td>1.10</td>
<td>1.75</td>
<td>-0.70</td>
<td>0.36</td>
<td>0.93</td>
<td>0.17</td>
<td>-0.13</td>
</tr>
<tr>
<td>15</td>
<td>IMP</td>
<td>1.92</td>
<td>-0.62</td>
<td>0.36</td>
<td>1.31</td>
<td>1.58</td>
<td>-0.90</td>
<td>0.27</td>
<td>0.92</td>
<td>&lt;0.01*</td>
<td>0.26</td>
</tr>
<tr>
<td>16</td>
<td>IMP</td>
<td>2.51</td>
<td>-0.39</td>
<td>0.57</td>
<td>1.45</td>
<td>2.68</td>
<td>-0.26</td>
<td>0.62</td>
<td>1.52</td>
<td>0.72</td>
<td>-0.09</td>
</tr>
<tr>
<td>17</td>
<td>IMP</td>
<td>2.11</td>
<td>-0.55</td>
<td>0.32</td>
<td>1.38</td>
<td>2.30</td>
<td>-0.43</td>
<td>0.61</td>
<td>1.47</td>
<td>0.07</td>
<td>-0.19</td>
</tr>
<tr>
<td>18</td>
<td>IMP</td>
<td>3.09</td>
<td>-0.35</td>
<td>0.76</td>
<td>1.41</td>
<td>2.42</td>
<td>-0.42</td>
<td>0.62</td>
<td>1.44</td>
<td>0.24</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Note. IA = inattention; HYP = hyperactivity; IA = impulsivity; ESSD = expected score standardized difference; SIDS = signed item difference in sample; UIDS = unsigned item difference in sample.

*Significant after correcting for multiple comparisons.
Figure 1. Schematic representations of the two correlated models of ADHD symptoms. Both models allow for correlations among the factors based on literature and theory. IA = inattention; HI = hyperactivity/Impulsivity; HYP = hyperactivity; IMP = impulsivity.
Model 3. Two-Factor Bifactor Model

Model 4. Three-Factor Bifactor Model

Figure 2. Schematic representations of the two bifactor models of ADHD symptoms. Models 3 and 4 are orthogonal. ADHD = attention-deficit/hyperactivity disorder; IA = inattention; HI = hyperactivity/impulsivity; HYP = hyperactivity; IMP = impulsivity.
Figure 3. Correlated factor models of ADHD symptoms in college students. IA = inattention; HI = hyperactivity/impulsivity; HYP = hyperactivity; IMP = impulsivity. All path coefficients are standardized estimates and significant at $p < .001$. 

Model 1. Correlated Two-Factor Model

Model 2. Correlated Three-Factor Model
Figure 4. Two-factor bifactor model of ADHD symptoms in college students. ADHD = attention-deficit/hyperactivity disorder; IA = inattention; HI = hyperactivity/impulsivity; HYP = hyperactivity; IMP = impulsivity. All path coefficients are standardized estimates. All paths, except those in bold, were significant (p < .001).

*p = .002.
Figure 5. Item information functions for inattention items (Panel A) and the test information function (Panel B) and its standard error (Panel C) for the Inattention (IA) scale. For Panels A and B, high levels of information indicate high reliability of the scale at a given point along the latent trait.
Figure 6. Category response curves for inattention items. Each curve indicates the likelihood for a level of an item to be endorsed at a given level of the inattention latent trait. The first curve for each item is for a response of 1, the second curve is for a response of 2, and so forth.
Inattention 7 – Lose things

Inattention 8 – Easily

Inattention 9 – Forgetful

Figure 6 – continued.
A. Item information function curves for hyperactivity items

![Item information function curves for hyperactivity items](image1)

B. Test information function curve for HYP scale

![Test information function curve for HYP scale](image2)

C. Standard error of measurement plot for the HYP scale

![Standard error of measurement plot for the HYP scale](image3)

Figure 7. Item information functions for hyperactivity items (Panel A) and the test information function (Panel B) and its standard error (Panel C) for the Hyperactivity (HYP) scale. For Panels A and B, high levels of information indicate high reliability of the scale at a given point along the latent trait.
Figure 8. Category response curves for hyperactivity items. Each curve indicates the likelihood for a level of an item to be endorsed at a given level of the inattention latent trait. The first curve for each item is for a response of 1, the second curve is for a response of 2, and so forth.
A. Item information function curves for impulsivity items

![Item information function curves for impulsivity items](image)

B. Test information function curve for IMP scale

![Test information function curve for IMP scale](image)

C. Standard error of measurement plot for the IMP scale

![Standard error of measurement plot for the IMP scale](image)

Figure 9. Item information functions for impulsivity items (Panel A) and the test information function (Panel B) and its standard error (Panel C) for the Impulsivity (IMP) scale. For Panels A and B, high levels of information indicate high reliability of the scale at a given point along the latent trait.
Figure 10. Category response curves for impulsivity items. Each curve indicates the likelihood for a level of an item to be endorsed at a given level of the inattention latent trait. The first curve for each item is for a response of 1, the second curve is for a response of 2, and so forth.
The Florida State University
Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673 · FAX (850) 644-4392

APPROVAL MEMORANDUM

Date: 2/19/2016

To: Rebecca Lynch

Address: 1107 W. Call St. Tallahassee, FL 32306
Dept.: PSYCHOLOGY DEPARTMENT

From: Thomas L. Jacobson, Chair

Re: Use of Human Subjects in Research
Psychometric Properties of an Adult ADHD Rating Scale: The BAARS-IV

The application that you submitted to this office in regard to the use of human subjects in the proposal referenced above have been reviewed by the Secretary, the Chair, and one member of the Human Subjects Committee. Your project is determined to be Expedited per 45 CFR § 46.110(7) and has been approved by an expedited review process.

The Human Subjects Committee has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval does not replace any departmental or other approvals, which may be required.

If you submitted a proposed consent form with your application, the approved stamped consent form is attached to this approval notice. Only the stamped version of the consent form may be used in recruiting research subjects.

If the project has not been completed by 2/17/2017 you must request a renewal of approval for continuation of the project. As a courtesy, a renewal notice will be sent to you prior to your expiration date; however, it is your responsibility as the Principal Investigator to timely request renewal of your approval from the Committee.
You are advised that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report, in writing any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving human subjects in the department, and should review protocols as often as needed to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Human Research Protection. The Assurance Number is FWA00000168/IRB number IRB00000446.

Cc: Janet Kistner, Advisor
HSC No. 2016.17507
APPENDIX B

INSTITUTIONAL REVIEW BOARD
PROJECT RENEWAL MEMORANDUM

The Florida State University
Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673 · FAX (850) 644-4392

RE-APPROVAL MEMORANDUM

Date: 1/27/2017

To: Rebecca Lynch
Address: 1107 W. Call St. Tallahassee, FL 32306
Dept.: PSYCHOLOGY DEPARTMENT

From: Thomas L. Jacobson, Chair

Re: Re-approval of Use of Human Subjects in Research
Psychometric Properties of an Adult ADHD Rating Scale: The BAARS-IV

Your request to continue the research project listed above involving human subjects has been approved by the Human Subjects Committee. If your project has not been completed by 1/26/2018, you must request a renewal of approval for continuation of the project. As a courtesy, a renewal notice will be sent to you prior to your expiration date; however, it is your responsibility as the Principal Investigator to timely request renewal of your approval from the committee.

If you submitted a proposed consent form with your renewal request, the approved stamped consent form is attached to this re-approval notice. Only the stamped version of the consent form may be used in recruiting of research subjects. You are reminded that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report in writing, any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor are reminded of their responsibility for being informed concerning research projects involving human subjects in their department. They are advised to review the protocols as often as
necessary to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

Cc: Janet Kistner, Advisor
HSC No. 2017.20249
REFERENCES


Grenwald-Mayes, G. (2002). Relationship between current quality of life and family of origin dynamics for college students with Attention-Deficit/Hyperactivity Disorder. *Journal of Attention Disorders, 5*, 211–222.


BIOGRAPHICAL SKETCH

CURRICULUM VITAE

REBECCA J. (MAZZULLO) LYNCH, M.S.

EDUCATION

2016-Present  Psychology Resident, Clinical Psychology Predoctoral Internship (APA-Accredited)
University of Mississippi Medical Center/G.V. (Sonny) Montgomery VAMC Consortium, Jackson, MS

2012-Present  Doctoral Candidate, Clinical Psychology (APA-Accredited)
Florida State University, Tallahassee, FL
Dissertation Title: *The Psychometric Properties of the Barkley Adult ADHD Rating Scale—IV (BAARS-IV) in a College Sample*
Dissertation Prospectus Defended: 10/2015
Advisor: Janet A. Kistner, Ph.D.
Dissertation Committee: Janet A. Kistner, Ph.D. (Chair), Sara A. Hart, Ph.D., Michael J. Kofler, Ph.D., Frances A. Prevatt, Ph.D., Christopher Schatschneider, Ph.D.

2009-2012  Master of Science, Clinical Psychology (APA-Accredited)
Florida State University, Tallahassee, FL
Thesis Title: *The Impact of Perceptual Bias and Social Status in Predicting Childhood Aggression: A Replication Study*
Thesis Defended: 3/2012
Advisor: Janet Kistner, Ph.D.
Thesis Committee: Janet A. Kistner, Ph.D., Thomas E. Joiner, Ph.D., Christopher Schatschneider, Ph.D.

2003-2006  Bachelor of Arts, Psychology and Sociology
State University of New York at Buffalo, Buffalo, NY
*Magna cum laude with High Distinction in Psychology*

PEER- Reviewed Journal Publications


**Peer-reviewed abstracts**


MANUSCRIPTS IN PREPARATION


ORAL PRESENTATIONS


POSTER PRESENTATIONS


RESEARCH EXPERIENCE

2009-Present  Graduate Research Assistant, Child Development and Psychopathology Laboratory
Faculty Supervisor: Janet A. Kistner, Ph.D.
Department of Psychology, Florida State University
- Assist with database management, statistical analyses, literature searches, manuscript preparation for review at peer-reviewed journals, and dissemination of findings at national and international conferences.
- Supervise the execution of multiple research projects in the lab; coordinate cross-site collaborations; organize weekly laboratory meetings; assist with the hiring, supervision, and training of undergraduate and full-time research assistants.

2012-2016  Clinical Research Assistant, Adherence Intervention to Promote Optimal Use of Insulin Pumps in Adolescents with Type 1 Diabetes
Principal Investigator: Kimberly A. Driscoll, Ph.D.
NIH/NIDDK K23 DK091558
College of Medicine, Florida State University
- Assisted with an ongoing NIH-funded randomized clinical trial to improve metabolic control of Type I diabetes in an adolescent population, with a specific focus on utilizing technology to optimize adherence and health outcomes.
  - Administered psychosocial assessment measures; assessed insulin pump knowledge/skills and re-instructed participants on knowledge/skill deficits; downloaded insulin pump and examined data; managed diabetes patient appointments; supervised undergraduate research assistants; managed databases; recruited study participants.
  - Aided in manuscript preparation.

2009-2010  Graduate Research Assistant, Visual Information Processing Laboratory
Supervisors: Ralph Radach, Ph.D., Jessi C. Hill, Ph.D.
Department of Psychology, Florida State University
- Assisted with two studies using eye tracking technology and oculomotor tasks.
  - Administered a variety of eye tracking tasks to children with and without Attention-Deficit/Hyperactivity Disorder (ADHD) as part of a study designed to test the nature of the executive deficits found in ADHD as well as to test for differences that may exist between the subtypes on oculomotor tasks; collected parent report data; administered a brief intelligence test and a measure of word reading accuracy and fluency; trained and supervised undergraduate research assistants on the study procedures and assessments; assisted with data management and data analysis.
  - Utilized eye tracking technology in an elementary school to examine various aspects of reading behavior in grade-school children.
2008-2009  
**Research Assistant, Relational Aggression, Victimization, and Adjustment During Middle Childhood**  
Principal Investigator: Jamie Ostrov, Ph.D.  
NICHD R03 HD059781  
Department of Psychology, University at Buffalo  
- Assisted in projects concerning the role of anger and emotion regulation in the development of aggressive behavior in preschoolers.  
  - Conducted school-based behavioral observations, including recording detailed descriptions of physical aggression, relational aggression, and prosocial behaviors; administered brief developmentally-appropriate interviews with children.  
  - Attended weekly lab meetings, focusing on reading and discussing relevant scholarly articles.

2006-2009  
**Research Support Specialist, The Adolescent and Family Development Project/The Neighborhood Family Development Project**  
Principal Investigator: Craig Colder, Ph.D.  
NIDA R01 DA019631  
Department of Psychology, University at Buffalo  
- Assisted in two longitudinal projects funded by the National Institute of Drug Abuse examining the development of behavior problems and substance use initiation in childhood.  
  - Conducted research interviews in laboratory and home settings; administered multiple neuropsychological tasks and brief tests of intelligence; collected physiological response data.  
  - Assisted in the development of study protocols; managed and analyzed data using SAS statistical software; programmed computer tasks and questionnaires; trained interviewers on study protocol and assessments; participant scheduling.  
  - Attended weekly lab meetings focused on data analysis, project-related matters, and discussing relevant scholarly articles.

2005-2009  
**Research Assistant/Research Counselor, Cognitive Processes and Clinical Response in ADHD**  
Principal Investigator: Larry W. Hawk, Jr., Ph.D.  
NIMH R01 MH069434  
Center for Children and Families, University at Buffalo  
- Assisted with the implementation of a double-blind, placebo-controlled study funded by the National Institute of Mental Health examining the effects of methylphenidate on laboratory tasks measuring neurocognitive processes implicated in ADHD and clinical factors in children ages 9-12 with ADHD.  
  - Served as a camp counselor for three consecutive summers at the Summer Research Camp (SRC); observed, monitored, and recorded behavior of children; administered several cognitive tests; operated physiological equipment; entered data.
Recruited participants; conducted preliminary phone screenings and in-person screening appointments to determine eligibility of children to participate in the SRC; diagnostic assessment involved a structured computerized clinical interview with parents, administration of standardized ratings scales to both children and parents, and standard intelligence and achievement testing and brief tests of hearing and vision with children; scored, entered, and managed data entry (SPSS).

Involved in project development; assisted with preparing protocols and materials for submission to the IRB for research approval, manuscript preparation, collecting parent and teacher report forms, and training research assistants.

2005-2009  
**Research Assistant, Psychophysiology Laboratory, Department of Psychology, University at Buffalo**  
**Supervisor:** Larry W. Hawk, Jr. Ph.D.  
Department of Psychology, University at Buffalo  
- Assisted in several different projects concerning the effects of medication and performance-based incentives on cognitive function in children with ADHD, motivated attention and emotion using prepulse inhibition of acoustic startle in undergraduate students, and the role of cognition in nicotine dependence in adult smokers.
  - Administered cognitive tests, physiological tasks, and questionnaire measures to undergraduate students and adult smokers.
  - Attended and participated in weekly lab meetings, which included discussion of cognitive task development, data analysis, and reading relevant scholarly articles.

**OTHER PROFESSIONAL TRAINING/WORKSHOPS**

2017  
**Parent-Child Interaction Therapy**  
University of Mississippi Medical Center, Jackson, MS  
**Training conducted by:** Dustin Sarver, Ph.D. (PCIT Train the Trainer)

2016  
**Cognitive Processing Therapy for PTSD**  
University of Mississippi Medical Center/G.V. (Sonny) Montgomery Department of Veterans Affairs, Jackson, MS  
**Training conducted by:** Lauren Graves, Ph.D. (VA CPT Trainer)

2012  
**Write Winning Grant Proposals Seminar**  
Florida State University, Tallahassee, FL  
**Presenter:** John D. Robertson, Ph.D.
2011 The Techniques and Tools of Cognitive Behavioral Analysis System of Psychotherapy (CBASP) Training for Clinicians
Florida State University General Psychology Clinic, Tallahassee, FL
Training conducted by: Martina Belz, Ph.D. (Certified CBASP-trainer)

2010 Using Modern Regression Discontinuity Analysis to Measure Effects of Educational Interventions Workshop
The Fairmont, Washington, D.C.
Presenter: Howard S. Bloom, Ph.D.

CLINICAL EXPERIENCE

2016-Present Predoctoral Psychology Resident, University of Mississippi Medical Center
Rotations: Pediatric Hematology/Oncology, Pediatric Wellness and Weight Clinic, Child Inpatient Behavioral Unit, CARES Psychiatric Residential Facility, Harbor House Chemical Dependency Services, Center for the Advancement of Youth, General Psychology Clinic
Supervisors: Scott Coffey, Ph.D.; Thomas D. Elkin, Ph.D., Leilani Greening, Ph.D., Cynthia Karlson, Ph.D., Crystal Lim, Ph.D., Dustin E. Sarver, Ph.D., Monica Sutton, Ph.D., Daniel Williams, Ph.D.

2012-2016 Clinical Research Assistant, Private Pediatric Endocrinology Practices of Larry Deeb, M.D. and Nancy Wright, M.D.
College of Medicine, Florida State University
Supervisor: Kimberly A. Driscoll, Ph.D.

2011-2014 Psychological Trainee, Attention Disorders Specialty Clinic
Louise R. Goldhagen Multidisciplinary Evaluation and Consulting Center, Florida State University
Supervisors: Beverly Atkeson, Ph.D., Anne Cituk, Ph.D., Lauren Hutto, Ph.D., Ann Selvey, Ph.D.

2013 Adolescent Group Therapist, Migrant Education Program and Gadsden County Schools
Louise R. Goldhagen Multidisciplinary Evaluation and Consulting Center, Florida State University
Supervisor: Kristen Schmidt, Ph.D.

2011 Child and Adolescent Group Therapist, Children’s Medical Services
Louise R. Goldhagen Multidisciplinary Evaluation and Consulting Center, Florida State University
Supervisor: Kristen Schmidt, Ph.D.
2010-2012  
**Psychological Trainee, General Psychology Clinic**  
Department of Psychology, Florida State University  
Supervisors: Kimberly Driscoll, Ph.D., Thomas Joiner, Ph.D., Pamela Keel, Ph.D., Therese S. Kemper, Ph.D., Donald Kerr, Ph.D., Norman B. Schmidt, Ph.D.

---

**TEACHING & SUPERVISORY EXPERIENCE**

2017  
**Student Clinical Supervisor for Doctoral Practicum Students, Harbor House Chemical Dependency Services**  
University of Mississippi Medical Center  
Supervisor: Scott Coffey, Ph.D.

2009-2016  
**Graduate Supervisor, Directed Individual Study**  
Department of Psychology, Florida State University  
Faculty Supervisor: Janet A. Kistner, Ph.D.

2013  
**Guest Lecturer, Abnormal Psychology**  
Department of Psychology, Florida State University  
Instructor: Mark Licht, Ph.D.

---

**UNIVERSITY SERVICE**

2013  
**Colloquium Co-Organizer for Understanding Self-Regulation and its Role in Language and Literacy Development and Academic and Behavioral Outcomes**  
Department of Psychology, Florida State University  
Director: Christopher Lonigan, Ph.D.

2010  
**Colloquium Co-Organizer for The Development of Mathematics Skills from Preschool through Early Elementary School**  
Department of Psychology, Florida State University  
Director: Christopher Lonigan, Ph.D.

---

**FELLOWSHIPS, HONORS, AND AWARDS**

2013, 2014  
**Clinical Graduate Research Development Award Conference Presentation Grant**  
Florida State University ($600)

2013, 2014  
**Congress of Graduate Students Conference Presentation Grant**  
Florida State University ($200)

2009-2013  
**Predoctoral Interdisciplinary Research Training (PIRT) Program Fellowship**  
Program Director: Christopher Lonigan, Ph.D.  
Florida Center for Reading Research, Florida State University (Total: $120,000; four years of support, with additional tuition remission, health benefits, and conference travel support)
AD HOC REVIEWER

Child Psychiatry and Human Development
Clinical Psychology Review
Journal of Clinical Child and Adolescent Psychology

PROFESSIONAL MEMBERSHIPS

Society for Research in Child Development
Association for Behavioral and Cognitive Therapies
Society for Research on Educational Effectiveness
American Psychological Association
Institute of Education Sciences