Peer Reviewed

Title:
Optimizing assessment procedures for attention-deficit/hyperactivity disorder (ADHD)

Author:
Shemmassian, Shirag Kevork

Acceptance Date:
2014

Series:
UCLA Electronic Theses and Dissertations

Degree:
Ph.D., Psychology 0780UCLA

Advisor(s):
Lee, Steve S

Committee:
Baker, Bruce L, Bookheimer, Susan Y, Piacentini, John C, Stigler, James W

Permalink:
http://escholarship.org/uc/item/6gf098p9

Abstract:

Copyright Information:
All rights reserved unless otherwise indicated. Contact the author or original publisher for any necessary permissions. eScholarship is not the copyright owner for deposited works. Learn more at http://www.escholarship.org/help_copyright.html#reuse
UNIVERSITY OF CALIFORNIA

Los Angeles

Optimizing assessment procedures for attention-deficit/hyperactivity disorder (ADHD)

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Psychology

by

Shirag Kevork Shemmassian

2014
ABSTRACT OF THE DISSERTATION

Optimizing assessment procedures for attention-deficit/hyperactivity disorder (ADHD)

by

Shirag Kevork Shemmassian

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2014

Professor Steve S. Lee, Chair

Attention-deficit/hyperactivity disorder (ADHD) is a highly heritable childhood-onset disorder characterized by developmentally extreme and impairing levels of inattention and/or hyperactivity/impulsivity that disrupt critical domains of academic, social, behavioral, emotional, and neuropsychological development. Improved diagnostic procedures are likely to facilitate early identification and timely delivery of interventions and to potentially reduce the risk of negative outcomes. Most research on the assessment of ADHD has focused on the clinical utility of DSM-IV symptoms and related clinical and diagnostic instruments, as well as the optimal symptom thresholds for accurate diagnosis. Unfortunately, examination of optimal strategies to integrate multi-informant data (e.g., parents and teachers), as well as development and evaluation of alternative empirically-derived assessment strategies across contexts (e.g., clinical settings, nationally representative samples), has received considerably less attention. Moreover, there is limited evidence on the association of different assessment strategies with respect to multi-
domain impairment. The purpose of this investigation was to describe the utility of multiple ADHD assessment strategies in identifying functionally impaired children across multiple periods of development and in separate clinical and population-based samples. The first study compared multiple strategies for using single and multi-informant ADHD data in their prospective prediction of functional impairment. The second study aimed to develop alternative ADHD symptom algorithms for evaluation against the DSM-IV with respect to predictions of multi-domain impairment. The third study described the base rates and psychometric properties of ADHD symptoms in a nationally representative sample. In addition, this study compared the utility of ADHD symptom algorithms developed within clinical and population-based samples with respect to associations with academic and family functioning, as well as general health outcomes. The use of various symptom algorithms should consider the intended purpose of the assessment – that is, to rule in or rule out ADHD. Whereas more sensitive (i.e., inclusive) algorithms optimally identified individuals experiencing ADHD-related impairment in case-control samples, somewhat more specific (i.e., exclusive) algorithms were most useful for identifying impairment in a nationally representative sample. Finally, several non-DSM-IV algorithms outperformed the DSM-IV in identifying children exhibiting ADHD-related impairment. With additional validation, identifying and implementing alternative symptom algorithms for ADHD and other mental health conditions may improve treatment planning by allowing clinicians to target the most clinically significant symptoms.
The dissertation of Shirag Kevork Shemmassian is approved.

Bruce L. Baker
Susan Y. Bookheimer
John Carl Piacentini
James W. Stigler

Steve S. Lee, Committee Chair

University of California, Los Angeles
2014
DEDICATION

To God, for His unyielding grace.

To my family and friends, who provided love and support.

To my advisors, who taught me how to learn and think.

And to Willa, for her patience, kindness, and belief in me.
### TABLE OF CONTENTS

**CHAPTER 1: Clinical assessment of ADHD**

![Page 1](#)

**CHAPTER 2: The predictive utility of four methods of incorporating parent and teacher symptom ratings of ADHD for longitudinal outcomes**

![Page 7](#)

- **Method**
  - ![Page 11](#)
- **Results**
  - ![Page 17](#)
- **Discussion**
  - ![Page 19](#)

**CHAPTER 3: Optimal algorithms for ADHD: Comparative validity of DSM-IV and alternative empirically-derived approaches**

![Page 29](#)

- **Method**
  - ![Page 32](#)
- **Results**
  - ![Page 37](#)
- **Discussion**
  - ![Page 40](#)

**CHAPTER 4: Cross-validation and development of empirically-derived ADHD assessment strategies: Insights from the National Longitudinal Study of Adolescent Health (Add Health)**

![Page 54](#)

- **Method**
  - ![Page 58](#)
- **Results**
  - ![Page 62](#)
- **Discussion**
  - ![Page 64](#)

**REFERENCES**

![Page 78](#)
# LIST OF TABLES

## Chapter 2.

Table 2-1. Demographic and clinical characteristics of diagnostic groups……………………………………..25

Table 2-2. Number of individuals meeting ADHD symptom criteria according to each symptom algorithm………………………………………………………………………………………………………26

Table 2-3. Classification accuracy of each ADHD symptom algorithm……………………………………………………………27

Table 2-4. Associations of ADHD algorithms with functional impairment and broad psychopathology approximately two years later…………………………………………………28

## Chapter 3.

Table 3-1. Demographic and clinical characteristics of diagnostic groups………………………………45

Table 3-2. Total predictive power values (TPVs) for parent-rated ADHD symptoms……………….46

Table 3-3. Total predictive power values (TPVs) for teacher-rated ADHD symptoms……………..48

Table 3-4. Number of individuals meeting ADHD symptom criteria according to each symptom algorithm……………………………………………………………………………………………50

Table 3-5. Classification accuracy of each ADHD symptom algorithm………………………………51

Table 3-6. Associations of ADHD algorithms with functional impairment and broad psychopathology……………………………………………………………………………………52

## Chapter 4.

Table 4-1. Demographic and clinical characteristics of the Add Health Sample at Wave 1……70

Table 4-2. Base rates for ADHD symptoms………………………………………………………………………………….71

Table 4-3. Total predictive power values (TPVs) for ADHD symptoms……………………………………73

Table 4-4. Number of individuals meeting ADHD symptom criteria according to each symptom algorithm…………………………………………………………………………………………….75
Table 4-5. Classification accuracy of each ADHD symptom algorithm.........................76
Table 4-6. Associations of ADHD algorithms with functioning.................................77
ACKNOWLEDGEMENTS

The author would like to thank the co-authors for their contributions given that portions of this manuscript appear in the following articles.


This work was supported by The Paul & Daisy Soros Fellowship for New Americans and the National Science Foundation through a graduate research fellowship to Shirag K. Shemmassian.

In addition, the author would like to thank the National Institutes of Health for funding the UCLA ADHD and Development Lab (R03AA020186-01 to Steve S. Lee).
Shirag K. Shemmassian

EDUCATION
University of California, Los Angeles, Los Angeles, CA
C. Phil. in Psychology, November 2012
M.A. in Psychology, December 2010

Cornell University, College of Human Ecology, Ithaca, NY
B.S. in Human Development, May 2008

HONORS AND AWARDS
Leadership Education in Neurodevelopmental Disabilities (LEND) Fellowship
National Science Foundation (NSF) Graduate Research Fellowship
The Paul and Daisy Soros Fellowship for New Americans
Graduate Summer Research Mentorship Fellowship
Eugene V. Cota-Robles Fellowship

PEER-REVIEWED PUBLICATIONS


CONFERENCE PRESENTATIONS


Li, J. J., Shemmassian, S., & Lee, S. S. (2010, May). Anxiety moderates the association between response variability and ADHD. Poster presented at the 22nd annual meeting of the Association for Psychological Science, Boston, MA.
CHAPTER ONE: CLINICAL ASSESSMENT OF ADHD

Attention-deficit/hyperactivity disorder (ADHD) is a highly heritable (Faraone et al., 2005; Waldman & Gizer, 2006) childhood-onset disorder characterized by developmentally extreme and impairing levels of inattention and/or hyperactivity/impulsivity (American Psychiatric Association [APA], 2000). ADHD is the most prevalent childhood mental disorder with a worldwide prevalence of 5.3% (Polanczyk, de Lima, Horta, Biederman, and Rohde, 2007) and in the United States, approximately 60% of physician visits related to mental health involve attention problems (Hoagwood, Kelleher, Feil, & Comer, 2000). The clinical and public health significance of ADHD is suggested by its prospective prediction of negative adolescent and adult outcomes including social, academic, and occupational impairment, elevated comorbidity with mood and externalizing disorders, as well as learning disabilities and speech problems (Barkley, Murphy, & Fischer, 2008; Biederman, Petty, Evans, Small, & Faraone, 2010; Lee, Lahey, Owens, & Hinshaw, 2008; Schieve et al., 2012). Moreover, two recent independent meta-analyses reported that childhood ADHD prospectively predicted adolescent and adult substance abuse/dependence across all substance types (Charach, Yeung, Climans, & Lillie, 2011; Lee, Humphreys, Flory, Liu, & Glass, 2011). In addition to deficits in sustained attention and inhibitory control, children with ADHD frequently exhibit executive dysfunction, including response inhibition, vigilance, working memory, and planning (see Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005). Thus, childhood ADHD disrupts critical domains of academic, social, behavioral, emotional, and neuropsychological development.

Given its clinical significance, diagnostic procedures for ADHD must be reliable and valid. Although improved diagnostic procedures are likely to facilitate early identification, timely delivery of interventions, and potentially reduce the risk of negative outcomes, most
research has prioritized the basic psychopathology and treatment of ADHD (Barkley, 2006). With respect to improving evidence-based assessment of ADHD (e.g., Lahey et al., 1994), previous work largely reflects efforts to validate individual ADHD instruments (i.e., structured and semi-structured interviews, symptom ratings scales, and broadband rating scales; see Pelham, Fabiano, & Massetti, 2005). However, far less is known about how to optimally integrate multi-method/informant ADHD data and the comparative utility of categorical and dimensional diagnostic data (see Marcus & Barry, 2011; Owens & Hoza, 2003). Considering the short- and long-term effects of ADHD on academic, social, behavioral, and emotional functioning, the importance of improving diagnostic procedures is difficult to overstate.

**ADHD Symptomatology**

The Diagnostic and Statistical Manual of Mental Disorders – 4th edition, Text Revision (DSM-IV-TR; APA, 2000) characterizes ADHD according to nine symptoms of inattention and nine symptoms of hyperactivity/impulsivity. ADHD diagnostic criteria consist of: 1) six or more inattention and/or hyperactivity/impulsivity symptoms; 2) clinically significant impairment in social, academic, or occupational functioning in two or more settings (e.g., home and school); and 3) some symptoms evident prior to age 7. Although correlated, a two-factor solution of inattention and hyperactivity symptoms is suggested by numerous factor analytic studies (e.g., Bauermeister, Alegria, Bird, Rubio-Stipec, & Canino, 1992; Healey et al., 1993; Lahey et al., 1988). The symptom threshold for ADHD was established by examining the associations of different parent and teacher symptom counts with “independent” criteria such as global impairment, clinician-validated diagnosis, and test-retest reliability estimates. Some ADHD symptoms were empirically derived (Atkins, Pelham, & Licht, 1985) whereas others were based on parent and teacher report (e.g., Milich, Widiger, & Landau, 1987). Nevertheless, several
highly discriminating items (e.g., “often engages in physically dangerous activities without considering the possible consequences”) and those with high face validity (e.g., “often acts before thinking”) that were part of DSM-III or DSM-III-R were not included in DSM-IV (Lahey et al., 1994; Pelham, Gnagy, Greenslade, & Milich, 1992). Moreover, because symptoms were included and excluded based on a single sample of 380 4- to 17-year-old clinic-referred children (87% of the sample was 4-13 years-old; Lahey et al., 1994), it is unclear whether all ADHD symptoms are equally appropriate across development. That is, the DSM assumes that ADHD symptoms are developmentally invariant with respect to base rates, predictions of impairment, etc. in younger versus older children.

Current ADHD Assessment Strategies

Presently, there are several evidence-based assessment strategies for ADHD. Structured diagnostic interviews (e.g., Diagnostic Interview Schedule for Children, 4th edition, DISC-IV; Shaffer et al., 2000) address all relevant diagnostic criteria (e.g., age of onset, persistence) and are considered the gold standard based on their superior psychometric properties. It is important to note, however, that structured diagnostic interviews are psychometrically strong in part because interviews are standardized. Among 247 children, the DISC 2.3 demonstrated moderate to high inter-rater agreement between diagnoses derived from lay interviewers and clinicians (retest interval was 1–15 days for 83% of the sample; \( \kappa = .60 \); Schwab-Stone et al., 1996). Diagnostic agreement between DISC and clinician diagnoses was also acceptable (\( \kappa = .72 \)). Shaffer et al. (2000) reported strong test-retest reliability for ADHD diagnoses based on the DISC with the parent (\( \kappa = .79 \); mean retest interval = 6.6 days). Semi-structured interviews (e.g., Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Children- Present and Lifetime Version, KSADS-PL; Kaufman, Birmaher, Brent, Ryan, & Rao, 2000) also probe
diagnostic criteria and provide the interviewer with flexibility in querying key domains. Although the KSADS-PL is used extensively in clinical research with high inter-rater and test-retest reliability for internalizing and externalizing disorders, ADHD diagnoses from the KSADS-PL were slightly less reliable ($\kappa = .63$ for current versus .55 for lifetime diagnosis) than the DISC (Kaufman et al., 1997). Finally, in addition to children and adolescents, structured and semi-structured diagnostic interviews have good empirical support in the assessment of ADHD in young children (i.e., preschool) (Lee & Humphreys, 2011).

Rating scales are also often used in the assessment of ADHD because they are efficient and can be used with multiple informants. The Conners Rating Scales (CRS; Conners, Sitarenios, Parker, & Epstein, 1998a, 1998b) and the Achenbach System of Evidence Based Assessment (e.g., Child Behavior Checklist, CBCL; Achenbach & Rescorla, 2001) are administered conveniently and provide developmentally-sensitive and empirically-derived indicators of psychopathology based on large normative samples. Representative normative samples are crucial to the assessment of ADHD given that symptoms are common, particularly early in development. Second, despite lacking normative data, the Disruptive Behavior Disorder Rating Scale (DBD; Pelham et al., 1992) and the Swanson, Nolan, and Pelham, IV (SNAP-IV) (Swanson et al. 2001) use precise DSM-IV language for ADHD, oppositional defiant disorder (ODD), and conduct disorder (CD), thus providing a more direct assessment of DSM-IV phenomenology than the CBCL or CRS. Parent-rated DBD data from 116 children yielded coefficient alphas of .82 and .85 for the inattentive and hyperactive dimensions, respectively, with slightly higher estimates for teachers (Massetti, Pelham, & Gnagy, 2005). Test-retest and inter-rater reliability estimates for the DBD in that sample were excellent and DBD ratings were sensitive to diagnoses derived from the DISC. The SNAP-IV also has strong psychometric
properties. In a large sample of kindergarten through 5th grade children, parent SNAP-IV ratings yielded alphas of .90 and .79 for inattention and hyperactivity/impulsivity symptoms according to parents and .96 and .92, respectively, for the same domains according to teachers (Bussing et al., 2008). However, most ADHD rating scales prioritize the assessment of symptoms rather than diagnostic criteria (e.g., age of onset). Thus, formal diagnostic designations of ADHD typically cannot rely solely on rating scales.

Proposed Project

Most research on the assessment of ADHD has focused on the clinical utility of DSM-IV symptoms and related clinical and diagnostic instruments, as well as the optimal symptom thresholds for accurate diagnosis. Unfortunately, other important assessment considerations, including optimal strategies to integrate multi-informant data (e.g., parents and teachers), the comparative utility of diagnostic instruments based on categorical and dimensional data, as well as development and evaluation of alternative empirically-derived assessment strategies, have received considerably less attention (see Marcus & Barry, 2011; Owens & Hoza, 2003; Power, Costigan, Leff, Eiraldi, & Landau, 2001; Shemmassian & Lee, 2012; Shemmassian & Lee, 2014 for exceptions). Moreover, there is limited evidence on the association of different assessment strategies with respect to multi-domain impairment. Despite the DSM requirement that ADHD must be accompanied by significant social, academic, or occupational impairment, previous research has largely used global impairment to determine optimal symptom thresholds and instrument utilities (e.g., Pelham et al., 2005; Lahey et al., 1994). To improve traction on the association of ADHD assessment methods and strategies with functional impairments, more refined measures of impairment, across relevant domains (e.g., social, academic), must be further explored.
The goal of this dissertation is to describe the utility of multiple ADHD assessment strategies in identifying functionally-impaired children across multiple periods of development and in separate clinical and population-based samples. The specific aims are as follows:

1) To evaluate multiple strategies for using single and multi-informant ADHD data (e.g., parent, teacher, and combined) in a sample of 195 children (Wave 1; 98% of the sample was ages 6 to 9) evaluated through the ADHD and Development Lab at UCLA (Sample 1, PI: Lee). Specifically, multiple assessment strategies for ADHD will be directly compared based on their prospective prediction of functional impairment assessed at a two-year follow-up (Wave 2; 91% of the sample was 8-11 years-old).

2) To examine the psychometric properties of ADHD symptoms (e.g., utility estimates, positive and negative predictive power) among 84 7- to 12-year old children from Sample 1 followed prospectively for two years (Wave 2) and to develop alternative symptom algorithms for evaluation against DSM-IV with respect to predictions of multi-domain impairment.

3) To describe the psychometric properties of ADHD symptoms in the National Longitudinal Study of Adolescent Health (Add Health; Sample 2) and to test the utility of ADHD symptom algorithms identified from Sample 1 with respect to academic and family functioning, as well as general health outcomes.
CHAPTER TWO: THE PREDICTIVE UTILITY OF FOUR METHODS OF INCORPORATING PARENT AND TEACHER SYMPTOM RATINGS OF ADHD FOR LONGITUDINAL OUTCOMES

Although ADHD symptoms are often persistent (Biederman et al., 2010) and predictive of functional impairment (Lee et al., 2008), the long-term dysfunction is not equally evident in all ADHD youth. For example, in a 10-year longitudinal study of 215 boys with \( n=110 \) and without \( n=105 \) ADHD, 65% of children with ADHD at baseline no longer met full diagnostic criteria 10 years later (Biederman et al., 2010). Nevertheless, 78% of the ADHD group in the same study exhibited persistent ADHD symptoms that were associated with elevated comorbidity, as well as greater educational and social impairment. In another study, 71% of girls with childhood ADHD exhibited persistent ADHD at a 5-year follow-up (Mick et al., 2011). However, girls whose ADHD had remitted continued to exhibit multi-domain impairment relative to a non-ADHD comparison group. Further substantiating the notion of significant variability within ADHD youth with respect to later outcomes (e.g., impairment), an eight-year longitudinal study of 118 4- to 6-year-old children with ADHD (Lahey, Pelham, Loney, Lee, & Willcutt, 2005) reported a highly dispersed pattern of impairment among children at a given assessment. Finally, in the DSM-IV ADHD Field Trials, whereas inattention was more strongly associated with school impairment, hyperactivity was more strongly associated with home impairment (Lahey et al., 1994). Taken together, individuals with ADHD exhibit largely heterogeneous symptomatology and impairment relative to one another across time points and settings, necessitating further investigation of the situational validity of data for assessing child psychopathology and evaluating treatment planning (De Los Reyes & Kazdin, 2005). Thus, identifying the most valid methods for assessing childhood ADHD symptoms using multi-
informant data is a significant priority. Implications from this work include helping clinicians to determine which individuals are at greatest risk for continued impairment and thus, need treatment, efforts for which have thus far yielded discouraging results (Sayal et al., 2010).

The incremental utility of multi-informant ratings of ADHD reflects well-established informant discrepancies of child psychopathology (Achenbach, McConaughy, & Howell, 1987). Informant discrepancies have important clinical implications because they may lead to different conclusions, including the assessment of the client’s need for treatment (De Los Reyes & Kazdin, 2005). Despite the centrality of multi-informant data to the assessment of ADHD, relatively little work has addressed how these data should be integrated. That is, new knowledge that evaluates which informant or diagnostic algorithm best identifies impaired children is sorely needed (Achenbach, 2006). This is especially important given the well-established discrepancy between parent and teacher ratings of child psychopathology, as well as clinicians’ perceptions of the utility of multi-informant data. Mental health professionals perceived teachers as better informants of hyperactivity and inattention than mothers; children were not seen as useful informants (Loeber, Green, & Lahey, 1990). However, these perceptions, as well as the utility of combining multi-informant data, have rarely been evaluated empirically. For example, although combining multi-informant data to diagnose ADHD using the ‘or rule’ (i.e., symptom is present if endorsed by the parent or teacher; Piacentini, Cohen, & Cohen, 1992) is valid (Lahey et al., 1998), this approach increases the number of true positives, as well as the number of false positives, which may contribute to the diagnostic instability of ADHD. Classification rates of ADHD in another study varied dramatically depending on how the data were derived (Valo & Tannock, 2010): that is, when data were derived from one or two informants (i.e., parents and teachers), a semi-structured clinical interview and/or rating scale, or a certain symptom algorithm
(variations of the ‘or rule’ and ‘and rule’ [i.e., symptom is present if endorsed by the parent and teacher]), classification rates varied from as few as 52 cases (42%) to as many as 103 cases (84%). Additionally, there may also be an optimal number, or even specific symptoms, that optimally assess ADHD. Among 5- to 12-year-old children, the ADHD symptoms that best predicted ADHD diagnostic status were different for parents (i.e., “avoids tasks that require sustained mental effort” and “has difficulty sustaining attention in tasks or play activities”) versus teachers (“loses things necessary for tasks or activities” and “has difficulty organizing tasks and activities”) (Power et al., 2001). Therefore, unlike DSM-based approaches, which treat all informants and symptoms equivalently, combining symptoms according to their predictive accuracy (and perhaps informant) may constitute a more effective diagnostic strategy for ADHD (Power et al., 2001).

Recent data described practices to incorporate multi-informant ADHD assessment data. For example, in a study of 232 5- to 10-year-old children with ($n=121$) and without ($n=111$) ADHD, parent ratings of ADHD optimally identified globally impaired children, whereas teacher ratings of ADHD most accurately identified children negatively regarded by peers (Shemmassian & Lee, 2012). In addition, Vaughn & Hoza (2013) found that teacher ratings of ADHD symptoms significantly improved predictions of externalizing and internalizing behaviors relative to exclusive reliance on parent ratings among 267 7- to 11-year-old children with ($n=185$) and without ($n=82$) ADHD. However, a structured diagnostic interview with the parent did not significantly improve predictions of consensus diagnoses (by five clinical or school psychologists) beyond parent and teacher symptom ratings. Given the incremental validity of parent and teacher ratings overall and for ADHD in particular (Johnston & Murray, 2003), multi-informant ratings of ADHD are prioritized (Pelham et al., 2005). Despite strong evidence for this
recommendation, far less is known about the *predictive utility* of different assessment methods incorporating multi-informant data.

Overall, ADHD assessment is critical to delivering services and monitoring treatment outcomes. Thus, empirically evaluating parent and teacher ratings of ADHD, individually and collectively, with respect to their prediction of future clinically significant outcomes remains a priority. Thus, we rigorously ascertained 195 6- to 10-year-old children (Wave 1) with and without ADHD and again at a two-year follow-up (Wave 2; Specific Aim 1) using multiple informants and methods across psychopathology and functional impairment domains. Separate parent and teacher ratings of ADHD were obtained using identical versions of the DBD rating scale (Pelham et al., 1992), and four common ADHD algorithms were evaluated: (1) parent only; (2) teacher only; (3) parent *or* teacher (‘or rule’); and (4) parent *and* teacher (‘and rule’). Given that clinical significance requires functional impairment (Pelham et al., 2005), we compared the algorithms with respect to prospective predictions of broad psychopathology (e.g., Internalizing, Externalizing, and Total Problems) and multi-domain functional impairment (e.g., academic, social) that were unrelated to the determination of original ADHD caseness. As in the DSM-IV Field Trials (Lahey et al., 1994), we used academic and social impairment as outcomes to judge the validity of the ADHD algorithms. We also utilized broad dimensions of psychopathology as criteria for ADHD algorithms given their frequent comorbidity of ADHD in community-based and clinic-referred samples (Angold, Costello, Farmer, Burns, & Erkanli, 1999). We hypothesized that the ‘or rule’ algorithm, which is the most sensitive of the four ADHD symptom algorithms analyzed and incorporates parent and teacher ratings (Shemmassian & Lee, 2012), will maximally predict multiple dimensions of psychopathology as well as social and academic impairment two years later.
Method

Wave 1 Participants

Participants were evaluated through the ADHD and Development Lab at UCLA (Sample 1, PI: Lee) and consisted of 195 (70% male) ethnically diverse 6- to 10-year-old children ($M=7.4$, $SD=1.1$; 98% of the sample was 6–9 years) with ($n=104$) and without ($n=91$) ADHD according to the DISC. We recruited participants in this age range for several reasons. First, given the DSM-IV diagnostic requirement that some ADHD symptoms must be present before age 7, children were recruited in this age range to reduce potential recall bias. Additionally, identifying the most predictive symptom algorithms at these ages will facilitate recommendations for early screening and intervention programs. Finally, we wanted to minimize the potential influence of age censored samples by having a narrow rather than a highly dispersed age range, which may be vulnerable to non-random influences (e.g., puberty).

Ninety-two percent of Wave 1 participants self-reported their race-ethnicity as follows: 53% White/Caucasian, 7% as Black/African American, 10% as Latino/Hispanic, 4% as Asian, 22% as Biracial, and 4% Other. Both parent and child were required to be fluent in English and children were required to live with at least one biological parent no less than half time. Exclusionary criteria included a Full Scale IQ<70, as well as a previous diagnosis of a pervasive developmental, seizure, or neurological disorder that prevented full participation in the study. Participants were diagnosed with ADHD ($n=104$) if they met full diagnostic criteria for any subtype of ADHD according to the Diagnostic Interview Schedule for Children – Fourth Edition (DISC-IV; Shaffer et al., 2000). 42% ($n=44$) met diagnostic criteria for the Predominantly Inattentive Type, 12% ($n=12$) for the Predominantly Hyperactive Type, and 46% ($n=48$) for the Combined Type. Non-ADHD comparison youth were recruited, screened, and assessed
identically to probands. To improve the external validity of probands, youth who met criteria for any disorder other than ADHD (e.g., anxiety) were placed in the non-ADHD comparison group, thus conservatively increasing the similarity between ADHD probands and the comparison group (see Hinshaw, 2002; Lee et al., 2008 for similar procedures). The ADHD and non-ADHD comparison groups were comparable with regard to age, sex, and race-ethnicity. However, as expected, ADHD youth at Wave 1 exhibited more oppositional defiant disorder (ODD) and conduct disorder (CD) symptoms, were less academically competent, and had greater Social, Internalizing, Externalizing, and Total Problems than non-ADHD youth (see Table 2-1).

Wave 2 Participants

Approximately two years later, 92% of the original Wave 1 sample participated in a follow-up assessment when youth were 7- to 13-years-old (M=9.7, SD=1.3; 89% of the sample was 8–11 years). Ninety-two percent of Wave 2 participants self-reported their race-ethnicity as follows: 47% White/Caucasian, 7% Black/African American, 10% Latino/Hispanic, 4% Asian, 24% Biracial, and 8% Other. The retained participants (i.e., the 92% of the original Wave 1 participants that were assessed at Wave 2) did not differ from those that were not assessed at Wave 2 with respect to age (t=.35, p=.73), sex (χ²=.39, p=.55), race-ethnicity (χ²=1.2, p=.30), or ADHD (t=.28, p=.78), ODD (t=.60, p=.55), and CD (t=.93, p=.36) symptoms.

Procedures

At baseline (i.e., Wave 1), high-risk youth were recruited through talks at self-help groups for ADHD and study fliers distributed to local mental health service providers. Language in recruitment material specifically targeted youth with elevated levels of inattention and hyperactivity (but did not require a previous diagnosis of ADHD). Non-ADHD comparison youth were recruited from local elementary schools and pediatric offices using fliers containing
“neutral” language (i.e., did not refer to ADHD-related problems). Families who contacted our research office completed an initial telephone screening to determine preliminary study eligibility. For eligible families who satisfied the inclusion and exclusion criteria described above, parents were mailed rating scales and families were invited to our research lab for in-person assessments of child behavior and family functioning. Each child’s primary teacher was asked to complete parallel rating scales of child behavior. Approximately 15% of children were assessed while taking psychotropic medication during their laboratory visits (mostly stimulants). However, whenever possible, all parents and teachers were asked to provide ratings based on the children’s unmedicated behavior if they normally received medication. Similar procedures have been used in other ADHD studies (Lee et al., 2008). Parents consented and children assented to all relevant study procedures, which were approved by the IRB. All families were compensated $50 for their participation at Wave 1. They additionally received a written summary of selected results of the child’s cognitive and academic functioning, as well as psychopathology.

Approximately two years later (i.e., Wave 2), families were invited to our research lab for a follow-up in-person assessment of child behavior and family functioning. Assessment procedures at Wave 2 were highly parallel to those at Wave 1. Each child’s primary teacher was asked to complete parallel rating scales of child behavior. Parents consented and children assented to all study procedures, which were approved by the IRB. All families received $75 for their participation at Wave 2, and were provided written reports of the assessment upon request.

**Wave 1 Baseline Measures**

*Diagnostic Interview Schedule for Children - Fourth Edition (DISC-IV; Shaffer et al., 2000)*. This fully structured diagnostic interview with the parent probed full DSM-IV criteria for ADHD (e.g., symptoms, duration, age of onset, settings, and multi-domain impairment). Test–
retest reliability for ADHD from the DISC ranged from .51 and .64 in the DSM-IV Field Trials (Lahey et al., 1994). Diagnostic designations from the DISC have shown predictive validity in previous studies of ADHD, including predictions of child psychopathology as well as social and academic impairment (Lee et al., 2008; Owens, Hinshaw, Lee, & Lahey, 2009). The DISC-IV was administered at Waves 1 and 2, but only Wave 1 data were used for the current study. The Cronbach alpha was .88 for the number of Wave 1 inattention and hyperactivity symptoms in the current sample.

*Disruptive Behavior Disorder Rating Scale (DBD; Pelham et al., 1992).* Parents and teachers completed identical versions of the DBD and rated each ADHD symptom as “not at all,” “just a little,” “pretty much,” or “very much.” Symptoms rated as “pretty much” or “very much” were considered present (Lahey et al., 1998; Pelham et al., 1992). The DBD has excellent psychometric properties and is considered an evidence-based instrument in the assessment of ADHD (Pelham et al., 2005). Parent-rated inattention and hyperactivity symptoms at Wave 1 yielded Cronbach alphas of .94 and .91, respectively, and .94 for teacher-rated inattention and hyperactivity symptom counts.

*Two-year Follow-up Measures (i.e., Wave 2)*

*Child Behavior Checklist (CBCL) and Teacher Report Form (TRF) (Achenbach & Rescorla, 2001).* The CBCL and TRF are parallel parent- and teacher-reported measures, respectively, that provide developmentally-sensitive and empirically-derived indicators of psychopathology and functioning. Reliability estimates of the subscales used to assess functioning were as follows (parent, teacher): School (Academic) (.90, .93); Social Problems (.90, .95); Internalizing Problems (.91, .86); Externalizing Problems (.92, .89); Total Problems (.94, .95) (Achenbach & Rescorla, 2001). To reduce Type I error and to minimize inflated
associations secondary to shared method variance (e.g., parent ratings of ADHD predicting parent ratings of impairment), parent- and teacher-rated T-scores within each domain were averaged to create five composite measures of psychopathology and functional impairment.

When teacher data were missing (n=100), parent-rated T-scores were used exclusively. Crucially, however, parent- and teacher-rated T-scores within each domain were significantly correlated (Academic: $r=.69, p<.01$; Social Problems: $r=.38, p<.01$; Internalizing Problems: $r=.26, p=.01$; Externalizing Problems: $r=.35, p<.01$; Total Problems: $r=.43, p<.01$) (2-tailed). Moreover, the participants with missing Wave 2 teacher DBD data did not differ from those with available data with respect to age ($t=1.19, p=.24$), sex ($\chi^2=1.6, p=.25$), race-ethnicity ($\chi^2=2.0, p=.17$), or ADHD ($t=1.5, p=.51$), ODD ($t=.51, p=.80$), and CD ($t=.75, p=.06$) symptoms.

**Data Analyses**

Based on Wave 1 parent and teacher DBD Rating Scale data, children were considered as having met symptom criteria for ADHD according to four different methods: (A) parent report only (i.e., children who exceeded 6 or more symptoms of inattention or hyperactivity/impulsivity according to the parent); (B) teacher report only (i.e., children who met symptom criteria for ADHD according to the teacher); (C) parent and teacher ratings combined using the ‘or rule’ (Piacentini et al., 1992); and (D) combined parent and teacher ratings using the ‘and rule’ (i.e., symptom was present if endorsed by the parent and teacher). Meeting symptom criteria was coded as a binary variable for each algorithm (i.e., met criteria or not). One-hundred percent of parent data and 64% of teacher data were available at Wave 1. Because these person-centered analyses contrasted children who did versus did not meet symptom criteria for ADHD, complete data were a priority. Therefore, missing Wave 1 teacher DBD data were imputed using ten iterations of Markov Chain Monte Carlo in SAS PROC MI (Yuan, 2002). To approximate the
actual covariance structure, analyses of the four diagnostic methods were conducted using the mean of the ten separate imputations. The participants with missing Wave 1 teacher DBD data did not differ from those with available data with respect to age ($t=1.49, p=.14$), sex ($\chi^2=.01$, $p=.94$), race-ethnicity ($\chi^2=.06, p=.94$), or ADHD ($t=.75, p=.46$), ODD ($t=.38, p=.71$), and CD ($t=1.31, p=.19$) symptoms. The number of youth that met or did not meet symptom criteria according to each algorithm is presented in Table 2-2. We also calculated the sensitivity (i.e., correctly identified positive status), specificity (i.e., correctly identified negative status), positive predictive value (PPV; proportion of positive test results that are true positives), and negative predictive value (NPV; proportion of negative test results that are true negatives) of each symptom algorithm based on their correspondence with Wave 1 ADHD diagnostic status from the DISC-IV (Table 2-3).

Because the four ADHD symptom algorithms at Wave 1 produced different rates of ADHD (Shemmassian & Lee, 2012), we directly compared differences in the predictive validity of each Wave 1 ADHD algorithm with respect to broad measures of psychopathology (Internalizing, Externalizing, and Total problems), as well as normed ratings of academic and social functioning, from the CBCL and TRF at Wave 2. Specifically, we performed five separate multiple regression analyses for each of the four symptom algorithms controlling for age, sex, and the number of Wave 1 DISC-IV ODD symptoms. To improve the specificity of predictions from Wave 1 ADHD algorithms, we controlled for youth sex and age as well as the number of Wave 1 ODD symptoms given their incremental prediction of outcome and impairment beyond ADHD (Hinshaw, Lahey, & Hart, 1993). We then interpreted the Akaike information criterion ($AIC$; Akaike, 1974) of statistically significant regression coefficients to determine which algorithm was a better fit the data (lower values correspond to a better fitting model (Table 2-4).
AIC was used to determine the best fitting model because the symptom algorithms consisted of nested models (e.g., the parent only and teacher only algorithms are nested within the ‘or rule’ and ‘and rule’ algorithms). Bonferroni corrections were applied (i.e., for each predicted outcome domain; \( \alpha = 0.05/4 = 0.0125 \)) to offset multiple comparisons.

Results

**ADHD Symptom Algorithms: Convergence with ADHD Caseness at Wave 1**

Each of the four separate ADHD symptom algorithms (parent only, teacher only, ‘or rule,’ ‘and rule’; see Method section) was judged with respect to its correspondence with a positive ADHD diagnosis at Wave 1 according to the DISC-IV (Table 2-3). The parent only algorithm yielded the second greatest sensitivity (.73), specificity (.93), and NPV (.75) in addition to greatest PPV along with the ‘and rule’ (.93). Alternatively, the teacher only algorithm yielded the second-lowest sensitivity (.48), specificity (.70), and NPV (.54) along with the lowest PPV (.65). The ‘or rule’ algorithm, which required either informants’ symptom endorsement, yielded the greatest sensitivity (.88) and NPV (.83), but lowest specificity (.63) and second-lowest PPV (.73). Finally, the ‘and rule’ algorithm, which required both informants’ symptom endorsement, yielded the greatest specificity (.98) and PPV (.93; same as the parent only algorithm), but the lowest sensitivity (.25) and NPV (.53).

**Wave 1 ADHD Algorithms: Predictive Utility of Wave 2 Psychopathology and Impairment**

We compared the predictive utility of each of the four Wave 1 ADHD symptom algorithms, derived from parent and teacher ratings on the DBD rating scale. That is, separate multiple regressions compared predictions of Wave 2 psychopathology and functional impairment (derived from the CBCL and TRF) from each Wave 1 ADHD algorithm. All of the Wave 2 outcomes (i.e., Internalizing, Externalizing, and Total Problems, Academic, Social
Problems) consisted of the mean of the CBCL and TRF T-scores, although the CBCL T-Score was used exclusively when TRF data were missing (n=100) (Table 2-4).

Internalizing Problems. The Wave 1 ADHD ‘or rule’ algorithm was the best fitting model for predicting Wave 2 Internalizing Problems ($\beta=.30, p<.0125, AIC=1425$). The parent only and teacher only algorithms also significantly predicted Internalizing Problems ($\beta=.21, p<.0125, AIC=1435$ and $\beta=.13, p<.0125, AIC=1440$, respectively) (lower AIC values correspond to better model fit). The ‘and rule’ algorithm did not provide a good fit to Internalizing Problems ($\beta=-.01, p>.05, AIC=1443$).

Externalizing Problems. The Wave 1 ADHD ‘or rule’ algorithm optimally predicted Wave 2 Externalizing Problems ($\beta=.28, p<.0125, AIC=1343$). Once again, the teacher only ($\beta=.22, p<.0125, AIC=1349$) and parent only ($\beta=.22, p<.0125, AIC=1351$) algorithms also significantly predicted Externalizing Problems (lower AIC values correspond to better model fit). However, the ‘and rule’ did not significantly predict Externalizing Problems ($\beta=.16, p>.0125, AIC=1356$).

Total Problems. The Wave 1 ADHD ‘or rule’ algorithm ($\beta=.42, p<.0125, AIC=1393$) provided the best fitting model for predicting Wave 2 Total Problems. All other algorithms also significantly predicted Total Problems in the following order of model fit (lower AIC values correspond to better model fit): parent only ($\beta=.35, p<.0125, AIC=1408$); teacher only ($\beta=.22, p<.0125, AIC=1423$); and ‘and rule’ ($\beta=.17, p<.0125, AIC=1428$).

Academic. The Wave 1 ADHD ‘or rule’ algorithm optimally predicted Wave 2 academic functioning ($\beta=-.36, p<.0125, AIC=1380$). The remaining algorithms also significantly predicted Wave 2 academic functioning in the following order of model fit (lower AIC values correspond
to better model fit): teacher only ($\beta=-.24, p<.0125, AIC=1393$); parent only ($\beta=-.24, p<.0125, AIC=1394$); and ‘and rule’ ($\beta=-.20, p<.0125, AIC=1396$).

**Social Problems.** The Wave 1 ADHD ‘or rule’ algorithm once again was the best fitting model for predicting Wave 2 Social Problems ($\beta=.32, p<.0125, AIC=1299$). The parent only ($\beta=.28, p<.0125, AIC=1303$) and ‘and rule’ algorithms ($\beta=.20, p<.0125, AIC=1311$) also significantly predicted Social Problems (lower $AIC$ values correspond to better model fit), although the teacher only algorithm was a poor fit to Wave 2 Social Problems ($\beta=.14, p>.05, AIC=1315$).

**Discussion**

Few studies have directly examined the incremental validity and clinical utility of multi-informant ADHD assessment (Shemmassian & Lee, 2012; Vaughn & Hoza, 2013), and even fewer have tested their prospective prediction of psychopathology and impairment. To that end, we empirically evaluated the psychometric properties (i.e., classification accuracy) of four ADHD symptom algorithms in a clinical sample: (1) parent only; (2) teacher only; (3) parent or teacher (‘or rule’); and (4) parent and teacher (‘and rule’). The classification accuracy of the four symptom algorithms varied considerably, with the parent only algorithm effectively balancing (i.e., yielded psychometric estimates between the superior and inferior algorithms) sensitivity, specificity, and NPV, and yielding the superior PPV along with the ‘and rule’. In addition, based on an ethnically diverse sample of 195 children with ($n=104$) and without ($n=91$) ADHD, we performed multiple regressions to assess how well each ADHD algorithm at baseline (i.e., Wave 1) predicted independent measures of broad psychopathology (Internalizing, Externalizing, and Total problems scales), as well as academic and social functioning, from the CBCL and TRF at a two-year follow-up (i.e., Wave 2). Overall, the Wave 1 ADHD ‘or rule’ algorithm optimally
predicted multiple dimensions of Wave 2 psychopathology and functional impairment and was superior relative to the other algorithms (i.e., parent only, teacher only, ‘and rule’).

Given that the ‘or rule’ algorithm, based on combining parent and teacher ratings of ADHD symptoms, optimally predicted subsequent psychopathology and impairment outcomes, these results converge with previous studies that multi-informant data provide incremental validity in the assessment of ADHD (Johnston & Murray, 2003; Pelham et al., 2005; Tripp, Schaugency, & Clarke, 2006; Vaughn & Hoza, 2013). Nevertheless, these results provide support for incorporating multi-informant using the more sensitive ‘or rule’ approach only (and not the ‘and rule’). The consistency of our results with previous studies likely partly reflects the low to moderate agreement among informants on major dimensions of child psychopathology (e.g., Achenbach, 1987; Kolko & Kazdin, 1993). At Wave 1, only 60% of children were similarly identified as having ADHD according to separate parent and teacher ratings. These informant discrepancies are likely to partially reflect the contextual specificity in which behaviors are observed (e.g., parents more routinely observe children’s mood than teachers) (De Los Reyes, 2011). A recent study of 185 children with and 82 children without ADHD evaluated the incremental validity of parent and teacher ratings of ADHD in correctly classifying consensus clinician diagnoses based on comprehensive assessment data (i.e., behavior rating scales, tests of achievement, etc.) (Vaughn & Hoza, 2013). Parent and teacher ADHD ratings collectively explained virtually all of the variance and although parent ratings alone correctly classified ADHD, teacher ratings significantly improved model fit. Given the independent association of parent and teacher ratings with respect to designations of ADHD (Vaughn & Hoza, 2013), and unique situational validity of their ratings (De Los Reyes & Kazdin, 2005), sensitive approaches for incorporating multi-informant data in the assessment of ADHD remain strongly
indicated. We note that the parent only algorithm also predicted psychopathology and functional impairment, and therefore is a valid approach for ADHD assessment in the absence of teacher data. Nevertheless, although outcome measures averaged parent- and teacher-rated behaviors and functioning, parent ratings were used exclusively when teacher outcome data were missing \((n=100)\). Thus, although parent data were featured prominently in the outcomes, the ‘or rule’ consistently yielded the best fitting model for predicting psychopathology and impairment.

As expected, classification accuracy estimates of the four ADHD symptom algorithms with respect to ADHD diagnostic status derived from the DISC varied considerably, reflecting unique attributes of each method. That is, each algorithm was distinctive based on its superior sensitivity (‘or rule’; .88), specificity (‘and rule’; .98), PPV (parent only and ‘and rule’; .93) and NPV (‘or rule’; .83), as well as modest sensitivity (‘and rule’; .25), specificity (‘or rule’; .63), PPV (teacher only; .65) and NPV (‘and rule’; .53). Because the parent only algorithm yielded consistently strong psychometric properties, this may reflect the fact that ADHD caseness was based exclusively on the parent-completed DISC-IV (e.g., Achenbach, 1987); thus, this finding should be interpreted cautiously. Nevertheless, these preliminary data suggest that use of the different algorithms should consider the intended purpose of the assessment – that is, to rule in or rule out ADHD (Power et al., 2001). Whereas more specific algorithms may be the most useful for screening purposes (e.g., pediatric offices), more sensitive algorithms, such as the ‘or rule’, may be more appropriate for identifying individuals experiencing ADHD and related impairment (e.g., clinical settings, case-control research). Highly sensitive algorithms for screening purposes increase the number of false positives (i.e., Type 1 error) and may expose children to unnecessary treatments (e.g., psychotropic medication). Because case-control studies often include non-ADHD youth with sub-clinical levels of ADHD or other psychopathology (e.g.,
Hinshaw, 2002; Lee et al., 2008), these children may be more impaired than non-ADHD youth in population-based studies. Thus, rigorous diagnostic procedures must address full diagnostic criteria (e.g., duration, age of onset, settings, and multi-domain impairment) regardless of the informant(s) used. Additionally, diagnostic procedures must adequately predict important functional domains (Lahey et al., 1994) to evaluate need for treatment. Although symptoms are necessary to diagnostic procedures, symptoms lack precision because they are independent of impairment required for making diagnoses. For ADHD, informants may collectively endorse 6 or more unique symptoms using the ‘or rule’ despite individually endorsing a sub threshold number of symptoms (e.g., 5 symptoms of inattention and 5 symptoms of hyperactivity). Nevertheless, sub threshold youth may still be impaired and in need of treatment (Lahey et al., 2005).

Because childhood ADHD prospectively predicts numerous negative adolescent and adult outcomes (e.g., Biederman, Petty, Clarke, Lomedico, & Faraone, 2011; Lee et al., 2008), efforts to improve the validity of ADHD assessment must be prioritized. The reliable and accurate ascertainment of ADHD and related impairments will facilitate timely delivery of interventions and potentially reduce the risk of negative outcomes, especially given that impairment is the primary reason for treatment referrals and mediates predictions of ADHD-related negative outcomes (Angold et al., 1999; Pelham et al. 2005). A heuristic example of the potential value of this approach comes from the autism literature, where early identification during sensitive periods facilitated the development of behavioral interventions (e.g., applied behavior analysis) that promoted language and social development while concurrently reducing repetitive behaviors (Granpeesheh, Tarbox, & Dixon, 2009). Unfortunately, early school-based screening and intervention programs for ADHD are not uniformly effective: in a population-based study of 4 and 5-year-old children in England, none of the interventions provided [(1)
identification of children with elevated inattention/hyperactivity; (2) providing teachers with a book containing information about ADHD, as well as evidence-based teaching advice and classroom management; or (3) both] outperformed the “no intervention” condition for reducing ADHD symptoms or impairment (Sayal et al., 2010). Additional research of early screening and intervention programs using more active and evidence-based treatments for ADHD (e.g., parent training, behavior management; Pelham & Fabiano, 2008) is required to appropriately evaluate the effectiveness of such approaches, and with multiple samples and settings.

We note several important limitations of this study. First, the use of parent and teacher ratings to ascertain ADHD, as well as broad psychopathology and functioning, may have inflated associations through shared method and source variance (Campbell & Fiske, 1959; Gomez, Burns, Walsh, & de Moura, 2003). Although we attempted to minimize this influence by combining Wave 2 parent and teacher ratings of functional impairment, parent ratings replaced missing teacher data, which may have inflated associations. Validating the multi-informant symptom algorithms, such as those used in the current study, with additional impairment measures (e.g., peer sociometrics) is an important future direction. Second, although we implemented “state of the art” methods to impute missing Wave 1 data (Schafer & Graham, 2002), imputed data may not accurately represent actual clinical data in our sample and in the population overall, potentially yielding less accurate classification accuracy estimates and regression coefficients. Nevertheless, even a single set of imputations demonstrates precise estimates of missing data (Rubin, 1987). Thus, our use of the mean of ten separate imputations was quite conservative and rigorous. Third, teacher data were missing for a large number of participants at Wave 2 (n=100). Parent data were used exclusively for these cases, which likely inflated associations due to shared source variance. The decision to impute missing teacher data
at Wave 1 but not Wave 2 was done for several reasons. Imputing missing teacher data at Wave 1 was necessary to ensure that the baseline sample was sufficiently large enough to evaluate the psychometric properties of the four ADHD algorithms. Moreover, given that we used other Wave 1 variables as covariates in our analyses (e.g., age, sex, ODD symptoms), imputing missing teacher data from these variables was appropriate. On the other hand, we did not incorporate any Wave 2 data in our analyses other than our outcome variables. Thus, imputing missing Wave 2 data from Wave 1 variables was not strongly indicated. The fourth limitation is our inclusion of clinic-referred youth, which may have potentially produced biased findings (Kopec & Esdaile, 1990). The use of population-based samples, which reflect different base rates of symptoms, is necessary to understand the performance of various assessment methods in more representative samples. Finally, the consistently lower classification accuracy values associated with individual teacher vs. parent ratings of ADHD may also reflect that ADHD caseness was based exclusively on the parent-completed DISC-IV (e.g., Achenbach, 1987).

The current study evaluated the predictive utility of four ADHD symptom algorithms based on multi-informant data with diverse measures of broad psychopathology and functional impairment approximately two years later. The most sensitive algorithm (i.e., requiring that a symptom be endorsed by a parent or teacher; ‘or rule’) based on parent and teacher symptom ratings best predicted Internalizing, Externalizing, and Total Problems, as well as academic and social functioning. Future research should examine the predictive utility of multi-informant assessment algorithms in childhood for identifying adolescents and adults experiencing impairment, as well as in multiple contexts (e.g., schools, population-based samples) to determine which individuals may benefit from early intervention in order to potentially reduce the long term negative outcomes associated with childhood ADHD.
Table 2-1. Demographic and clinical characteristics of diagnostic groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>ADHD (n = 104)</th>
<th>Comparison (n = 91)</th>
<th>t/χ²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>7.31 (1.15)</td>
<td>7.51 (1.10)</td>
<td>-1.22</td>
</tr>
<tr>
<td>Number of Boys (%)</td>
<td>76 (73%)</td>
<td>60 (66%)</td>
<td>1.17</td>
</tr>
<tr>
<td>Number of non-White children (%)¹</td>
<td>45 (47%)</td>
<td>39 (46%)</td>
<td>.00</td>
</tr>
<tr>
<td>Number of DISC-IV ADHD Symptoms</td>
<td>12.33 (3.21)</td>
<td>3.07 (2.86)</td>
<td>21.12*</td>
</tr>
<tr>
<td>Number of DISC-IV ODD Symptoms</td>
<td>3.35 (2.43)</td>
<td>.96 (1.65)</td>
<td>8.14*</td>
</tr>
<tr>
<td>Number of DISC-IV CD Symptoms</td>
<td>1.33 (1.60)</td>
<td>.49 (.72)</td>
<td>4.77*</td>
</tr>
<tr>
<td>Academic Functioning †</td>
<td>42.58 (9.68)</td>
<td>48.74 (8.64)</td>
<td>-4.60*</td>
</tr>
<tr>
<td>Social Problems †</td>
<td>59.52 (7.84)</td>
<td>53.69 (5.05)</td>
<td>6.26*</td>
</tr>
<tr>
<td>Internalizing Problems †</td>
<td>56.67 (9.25)</td>
<td>49.49 (9.19)</td>
<td>5.42*</td>
</tr>
<tr>
<td>Externalizing Problems †</td>
<td>56.70 (8.57)</td>
<td>47.80 (7.82)</td>
<td>7.53*</td>
</tr>
<tr>
<td>Total Problems †</td>
<td>60.49 (7.49)</td>
<td>48.79 (9.74)</td>
<td>9.30*</td>
</tr>
</tbody>
</table>

Note. DISC-IV = Diagnostic Interview Schedule for Children, 4th edition; ADHD = attention-deficit/hyperactivity disorder; ODD = oppositional defiant disorder; CD = conduct disorder; *p < .05.

¹Ethnicity data available for 179 children.

†Average of CBCL and TRF subscales’ T-scores.
Table 2-2. Number of individuals meeting ADHD symptom criteria according to each symptom algorithm

<table>
<thead>
<tr>
<th>Symptom Algorithm</th>
<th>Parent</th>
<th>Teacher</th>
<th>‘Or Rule’</th>
<th>‘And Rule’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
<td>Absent</td>
<td>Present</td>
<td>Absent</td>
</tr>
<tr>
<td>ADHD (n = 104)</td>
<td>76</td>
<td>28</td>
<td>50</td>
<td>54</td>
</tr>
<tr>
<td>Comparison (n = 91)</td>
<td>6</td>
<td>85</td>
<td>27</td>
<td>64</td>
</tr>
<tr>
<td>Total</td>
<td>82</td>
<td>113</td>
<td>77</td>
<td>118</td>
</tr>
</tbody>
</table>

*Note. ADHD = attention-deficit/hyperactivity disorder; Parent = parent DBD only; Teacher = teacher DBD only; ‘Or Rule’ = combined parent and teacher DBD data using ‘or rule’; ‘And Rule’ = combined parent and teacher DBD data using ‘and rule’.

*aSymptoms rated on the Disruptive Behavior Disorder Rating Scale.
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent</td>
<td>.73</td>
<td>.93</td>
<td>.93</td>
<td>.75</td>
</tr>
<tr>
<td>Teacher</td>
<td>.48</td>
<td>.70</td>
<td>.65</td>
<td>.54</td>
</tr>
<tr>
<td>‘Or Rule’</td>
<td>.88</td>
<td>.63</td>
<td>.73</td>
<td>.83</td>
</tr>
<tr>
<td>‘And Rule’</td>
<td>.25</td>
<td>.98</td>
<td>.93</td>
<td>.53</td>
</tr>
</tbody>
</table>

*Note. ADHD = attention-deficit/hyperactivity disorder; Parent = parent DBD only; Teacher = teacher DBD only; ‘Or Rule’ = combined parent and teacher DBD data using 'or rule'; ‘And Rule’ = combined parent and teacher DBD data using 'and rule'; PPV = positive predictive value; NPV = negative predictive value.

*Classification accuracy based on DISC-IV ADHD diagnoses.*

*Symptoms rated on the Disruptive Behavior Disorder Rating Scale.*
Table 2-4. Associations of ADHD algorithms with functional impairment and broad psychopathology approximately two years later

<table>
<thead>
<tr>
<th>Symptom Algorithm</th>
<th>Academic Functioning&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Social Problems&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Internalizing Problems&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Externalizing Problems&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Total Problems&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent</td>
<td>-0.24**</td>
<td>0.28**</td>
<td>0.21**</td>
<td>0.22**</td>
<td>0.35**</td>
</tr>
<tr>
<td>Teacher</td>
<td>-0.24**</td>
<td>0.14</td>
<td>0.13**</td>
<td>0.22**</td>
<td>0.22**</td>
</tr>
<tr>
<td>‘Or Rule’</td>
<td>-0.36**</td>
<td>0.32**</td>
<td>0.30**</td>
<td>0.28**</td>
<td>0.42**</td>
</tr>
<tr>
<td>‘And Rule’</td>
<td>-0.20**</td>
<td>0.20**</td>
<td>-0.01</td>
<td>0.16*</td>
<td>0.17**</td>
</tr>
</tbody>
</table>

Note. ADHD = attention-deficit/hyperactivity disorder; Parent = parent DBD only; Teacher = teacher DBD only; ‘Or Rule’ = combined parent and teacher DBD data using 'or rule'; ‘And Rule’ = combined parent and teacher DBD data using 'and rule'; AIC = Akaike information criterion.  
*<i>p < .05.</i>  
**<i>p < .0125 (α value based on Bonferroni correction).</i>  
<sup>a</sup>Average of CBCL and TRF subscales’ T-scores. CBCL T-scores used when teacher data were missing.  
<sup>†</sup>Lowest AIC value within a functional domain.
CHAPTER THREE: OPTIMAL ALGORITHMS FOR ADHD: COMPARATIVE VALIDITY OF DSM-IV AND ALTERNATIVE EMPIRICALLY-DERIVED APPROACHES

Previous research has identified optimal diagnostic thresholds for ADHD (Lahey et al., 1994; Power et al., 2001), but less is known about optimal procedures to integrate multi-method/informant ADHD data, as well as the comparative utility of diagnostic instruments based on categorical and dimensional data. However, a heuristic example of the potential value of this approach was reported in the literature on autism. A study of 1,630 Autism Diagnostic Observation Schedule (ADOS) assessments, widely considered the gold standard in the assessment of autism (Lord et al., 1989), was conducted with 14 month to 16 year-old children with and without an autism spectrum disorder (ASD) (Gotham, Risi, Pickles, & Lord, 2007). Using exploratory multi-factor item response analysis and logistic regression, the optimal weighting of multiple domains to differentiate children with and without ASD was evaluated. A revised diagnostic algorithm, consisting of combined scores from the newly derived Social Affect and Restricted, Repetitive Behaviors domains, yielded better predictive utility (i.e., classification accuracy) than the ADOS diagnostic algorithm based on the Communication and Social domains. Most notably, the revised algorithm improved specificity by as much as 31 percent and performed similarly well in an independent sample (Oosterling et al., 2010). Thus, ADHD assessment procedures may be similarly improved from innovations in combining multi-informant/-method data. Crucially, improving ADHD assessment procedures will optimize diagnostic accuracy and facilitate early identification and timely delivery of interventions.

Although there are three primary evidence-based instruments for ADHD assessment (i.e., structured and semi-structured interviews, symptom rating scales, and broadband rating scales), each approach has advantages and limitations (Pelham et al., 2005; Shemmassian & Lee, 2012).
Structured diagnostic interviews (e.g., Diagnostic Interview Schedule for Children, 4th edition [DISC-IV]; Shaffer et al., 2000) are strongly recommended (Lahey & Willcutt, 2002) by virtue of their superior psychometric properties including test-retest reliability and predictive validity (Shaffer et al., 2000). However, administration is lengthy and impractical with multiple informants, despite the centrality of multiple informants in the assessment of ADHD (Johnston & Murray, 2003; Pelham et al., 2005; Power et al., 1998). This is problematic given that parents and teacher ratings are influenced not only by child characteristics, but also the context in which they are observed (i.e., contextual validity; De Los Reyes, 2011); combined parent and teacher data are also superior to single-informant data in predictions of functional impairment (Owens & Hoza, 2003; Shemmassian & Lee, 2012). Beyond their problems in clinical settings (Pelham et al., 2005), diagnostic interviews inherently reflect the limitations of DSM-IV. For example, inattention and hyperactivity were differentially associated with ADHD diagnostic status in the DSM-IV Field Trials (Frick et al., 1994), but they are equally weighted. That is, DSM-IV aggregates psychometrically divergent items to evaluate ADHD. More advanced methods, such as item response theory (IRT), directly evaluate the informativeness and discrimination of items reflecting individual differences of a latent trait (i.e., ADHD). Among 1,475 school-aged children, some inattention symptoms optimally discriminated the latent ADHD trait for parents (e.g., “often fails to give close attention to details or makes careless mistakes in schoolwork, work, or other activities”) and teachers (e.g., “often runs about or climbs excessively in situations in which it is inappropriate”), whereas different hyperactivity symptoms discriminated the traits for parents (e.g., “often runs about or climbs excessively in situations in which it is inappropriate”) vs. teachers (e.g., “often blurts out answers before questions have been completed”) (Gomez, 2008). Similar methods applied to ADHD symptoms in preschool children
suggested that current symptoms were useful to diagnosing ADHD, but several symptoms were redundant (Purpura, Wilson, & Lonigan, 2010). Taken together, each symptom of ADHD is unlikely to equally reflect the latent ADHD trait. Therefore, the diagnostic accuracy of ADHD may improve with a better understanding of which symptoms, across informants and inattention-hyperactivity dimensions, optimally predict ADHD and functional impairment.

Rating scales are not only efficient; they have strong empirical support in the assessment of ADHD (Pelham et al., 2005). In fact, children with ADHD diagnosed according to structured interviews are nearly identical to children ascertained with rating scales (DuPaul, Power, McGoe, Ikeda, & Anastopoulos, 1998). Crucially, ADHD rating scales provide useful data to examine symptom utility estimates, which are typically estimated from an item’s positive predictive power (PPP), negative predictive power (NPP), as well as its sensitivity and specificity. Among 5- to 12-year-old children with (n = 41) and without (n = 71) ADHD, parent- and teacher-rated ADHD symptoms were markedly different with regard to predictive power, but were most predictive of ADHD caseness at the two highest severity levels (i.e., “pretty much” and “very much” vs. “not at all” and “just a little”) (Power et al., 2001). Moreover, when algorithms of parent and teacher ratings of ADHD were combined empirically according to PPP and NPP values, they were highly effective at ruling in or ruling out ADHD (Power et al., 2001). Additionally, the optimal number of items for predicting ADHD diagnostic status varied by informant and inattention vs. hyperactivity, with diagnostic efficiency achieved with as few as two symptoms. However, Power et al. (2001) did not compare the utility of these empirically-derived algorithms against the DSM-IV-based thresholds for predicting functional impairment. Therefore, it is unknown how these alternative strategies compare with DSM-based strategies, a crucial consideration given the modest correlation between ADHD symptoms and impairment.
(Fabiano et al., 2006). One exception is a recent case-control study of 151 6- to 9-year-old children with (n = 76) and without (n = 75) ADHD (Wave 1) that evaluated the utility of various parent- and teacher-rated symptom algorithms against the DSM-IV approach in their associations with functional impairment (Shemmassian & Lee, 2014). More sensitive approaches to ADHD classification better predicted functional impairment and broad psychopathology and consistently outperformed the DISC-IV. Future research must not only identify symptoms and the algorithms with the highest utility estimates, but also continue to examine their ability to identify functionally impaired children against the DSM-IV approach across age groups and contexts.

Overall, knowledge about how parent and teacher ratings of ADHD should be optimally weighted to identify impaired children will improve the clinical ascertainment of ADHD in various demographic groups. Using 84 7- to 12-year-old children with and without ADHD that were evaluated approximately two years after an initial assessment as part of a larger longitudinal study (Sample 1), symptom utility estimates were calculated for separate parent and teacher ratings of ADHD. These estimates served as the basis for creating empirically-derived algorithms that were directly compared against DSM-IV based designations in predictions of differentiated measures of functional impairment and broad psychopathology unrelated to ADHD caseness (Specific Aim 2).

Method

Participants

Participants were 84 (69% male) ethnically diverse 7- to 12-year-old children (M = 9.5 years, SD = 1.3; 91% of the sample was 8–11 years) with (n = 42) and without (n = 42) ADHD according to the DISC-IV administered at Wave 2 that were evaluated approximately two years after an initial assessment (Wave 1) as part of a larger longitudinal study. Participants were
originally recruited through presentations to self-help groups and advertisements mailed to local elementary schools, pediatric offices, and clinical service providers. Exclusionary criteria included a Full Scale IQ < 70, a previous diagnosis of a pervasive developmental, seizure, or neurological disorder, or any medical condition that prevented full participation in the study. Both parent and child were required to be fluent in English and children were required to live with at least one biological parent no less than half time. Although 92% of the original Wave 1 sample participated in the Wave 2 follow-up, only 56% of the original sample provided complete data across all the measures and informants in the current set of analyses. 45% of participants in the current study self-identified as White/Caucasian, 7% as Black/African American, 13% as Latino/Hispanic, 5% as Asian, 21% as Biracial, and 9% Other. At Wave 2, the ADHD and non-ADHD groups were comparable with regard to age, sex, race-ethnicity, conduct disorder (CD) symptoms, and internalizing problems. However, ADHD youth exhibited more oppositional defiant disorder (ODD) symptoms, were less academically competent, and had greater social, externalizing, and total problems than non-ADHD youth (see Table 3-1). All analyses in the current study are solely based on Wave 2 participants with complete data (n = 84).

Procedures

Approximately two years after participating in a baseline assessment (see page 16), eligible families who satisfied the inclusion and exclusion criteria described above were invited to our research lab for follow-up in-person assessments of child behavior and family functioning. Each child’s parent and primary teacher were also asked to complete parallel rating scales of child behavior. If a child normally received medication, their parents and teachers were asked to provide ratings based on the child’s unmedicated behavior. Similar procedures have been used in
other ADHD studies (Hinshaw, 2002; Lee et al., 2008). Parents consented and children assented to all relevant study procedures, which were approved by the IRB.

Measures

Diagnostic Interview Schedule for Children - Fourth Edition (DISC-IV; Shaffer et al., 2000). This fully structured diagnostic interview with the parent probed full DSM-IV criteria for ADHD (e.g., duration, age of onset, impairment). Test–retest reliability for ADHD from the DISC ranged from .51 and .64 in the DSM-IV Field Trials (Lahey et al., 1994). Diagnostic designations from the DISC have shown predictive validity in previous studies of ADHD (Owens et al., 2009; Lee et al., 2008). The Cronbach’s alpha of the Wave 2 participants was .88 for both the total number of inattention and hyperactivity symptoms.

Child Behavior Checklist (CBCL) and Teacher Report Form (TRF) (Achenbach & Rescorla, 2001). The CBCL and TRF are parallel parent- and teacher-reported measures, respectively, that provide developmentally-sensitive and empirically-derived indicators of psychopathology and functioning. Reliability estimates of the subscales used to assess functioning were as follows (parent, teacher): School (Academic) (.90, .93); Social Problems (.90, .95); Internalizing Problems (.91, .86); Externalizing Problems (.92, .89); Total Problems (.94, .95) (Achenbach & Rescorla, 2001). Because parent- and teacher-rated T-scores within each domain were significantly correlated (Academic: r = .77, p < .01; Social Problems: r = .39, p < .01; Internalizing Problems: r = .25, p = .02; Externalizing Problems: r = .37, p < .01; Total Problems: r = .45, p < .01) (2-tailed), T-scores from parent and teacher ratings within each domain were averaged to create five composite measures of functioning. This procedure additionally protected against inflated associations due to shared method variance (e.g., parent ratings of ADHD predicting parent ratings of impairment) and reduced Type I error.
Disruptive Behavior Disorder Rating Scale (DBD; Pelham et al., 1992). Parents and teachers completed identical versions of the DBD and rated each ADHD symptom as “not at all,” “just a little,” “pretty much,” or “very much.” Symptoms rated as “pretty much” or “very much” were considered present (Lahey et al., 1998; Pelham et al., 1992). The DBD has excellent psychometric properties and is considered an evidence-based instrument in the assessment of ADHD (Pelham et al., 2005). The total number of parent-rated inattention and hyperactivity symptoms in this sample yielded Cronbach’s alphas of .94 and .91, respectively, and .94 for both teacher-rated inattention and hyperactivity symptom counts.

Diagnostic Ascertainment of ADHD

At Wave 2, parents were interviewed using the DISC-IV and completed the modules for ADHD and other disorders (e.g., disruptive behavior, depression, anxiety). Comorbid disorders were permitted to improve the external validity of ADHD probands. Participants were diagnosed with (n = 42) ADHD if they met full diagnostic criteria for any subtype of ADHD according to the DISC. 43% (n = 18) met diagnostic criteria for the Predominantly Inattentive Type, 14% (n = 6) for the Predominantly Hyperactive/Impulsive Type, and 43% (n = 18) for the Combined Type.

Data Analyses

ADHD Symptom Severity Thresholds

We began by separately calculating the total predictive value (TPV; i.e., hit rate) of each level (i.e., “not at all,” “just a little,” “pretty much,” “very much”) of each parent- and teacher-rated ADHD symptom against ADHD versus non-ADHD status derived from the DISC-IV. For example, we calculated separate TPVs for parent and teacher ratings associated with three comparisons: (1) “not at all” versus “just a little,” “pretty much,” or “very much;” (2) “not at all” or “just a little” versus “pretty much” or “very much;” and (3) “not at all,” “just a little,” and
“pretty much” versus “very much.” The symptom level at which each parent- and teacher-rated item yielded the highest TPV for ADHD caseness was considered optimal (Tables 2.2 and 2.3).

**Symptom Algorithms and Classification Accuracy**

Next, we created three separate algorithms for parent- and teacher-rated ADHD symptoms: (1) Classic: ≥ 6 of 9 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “pretty much” or “very much”; (2) Observed: ≥ 6 of 9 inattention and/or hyperactivity/impulsivity symptoms endorsed at their highest TPV level; (3) Sensitive: a third algorithm was created given a recent study’s finding that more sensitive approaches to meeting symptom criteria may be most useful for identifying children experiencing functional impairment and broad psychopathology in case-control study designs (Shemmassian & Lee, 2014). This algorithm was defined by ≥ 6 of 9 inattention and/or hyperactivity/impulsivity symptoms endorsed as “just a little,” “pretty much” or “very much.” The highest TPV for most parent- and teacher-rated symptoms in the current study was evident when symptoms were endorsed at the “pretty much” or “very much” levels (67% and 56% of symptoms for parents and teachers, respectively; see Tables 2.2 and 2.3). The number of individuals that did or did not meet symptom criteria according to each algorithm is presented in Table 3-4. We also calculated the sensitivity, specificity, PPV, and NPV of each symptom algorithm based on their correspondence with ADHD diagnostic status ascertained from the DISC-IV (Table 3-5).

**Associations with Functional Impairment**

Although individual ADHD symptoms differentially predict ADHD at different levels for parents and teachers (Pelham et al., 2005), it is unclear whether the different combinations of symptom levels and informants are superior to structured diagnostic interviews (e.g., DISC-IV) in predictions of functional impairment. We directly compared the concurrent validity of
empirically-derived versus DISC-IV ADHD diagnoses with normed ratings of academic and social functioning, as well as broad measures of psychopathology from the CBCL and TRF. Specifically, we performed separate multiple regressions for each symptom algorithm (i.e., 7 regressions total within each functional domain or area of psychopathology) controlling for the number of DISC-IV ODD symptoms. Age and sex were not controlled given that CBCL/TRF T-scores are age- and sex-normed. We then interpreted the $R^2$ values of statistically significant regression coefficients to determine which algorithm explained the greatest amount of variance for each outcome variable (higher values correspond to more variance explained) (Table 3-6). Bonferroni corrections were applied to each set of regressions (i.e., for each functional domain or area of psychopathology; $\alpha=.05/7=.007$) to control for multiple comparisons.

Results

Clinical Utility of Each ADHD Symptom

12 of 18 (56%) parent-rated ADHD symptoms were most predictive of ADHD diagnostic status when endorsed as “pretty much” or “very much,” whereas five of the symptoms were most predictive when endorsed as “very much,” and one symptom was most predictive when endorsed as “just a little,” “pretty much” or “very much.” 10 of 18 (56%) teacher-rated ADHD symptoms were most predictive of diagnostic status when endorsed as “pretty much” or “very much.” Seven additional symptoms were most predictive when endorsed as “very much,” whereas one symptom was most predictive when endorsed as “just a little,” “pretty much” and “very much” (see Tables 2.2 and 2.3 for TPV values and cutoffs).

Classification Accuracy of ADHD Symptom Algorithms
The three symptom algorithms (Classic, Observed, Sensitive; see Method section) were separately applied to parent and teacher ratings to classify children as meeting symptom criteria for any ADHD subtype.

**Parent Algorithms.** The Sensitive algorithm, which relaxed the symptom threshold, yielded the greatest sensitivity (1.00) and NPV (1.00), but the lowest specificity (.59) and PPV (.68). The Observed algorithm, constructed from symptom thresholds based on each symptoms’ predictive utility, several of which were most predictive when endorsed at the “very much” level, yielded the greatest specificity (.98) and PPV (.95), but the lowest sensitivity (.46) and NPV (.66). The Classic algorithm yielded psychometric properties between the Observed and Sensitive algorithms’. Specifically, the Classic algorithm yielded sensitivity and specificity values of .71 and .95, respectively, and PPV and NPV values of .94 and .77, respectively (see Table 3-5).

**Teacher Algorithms.** Symptom algorithms based on teachers’ symptom ratings were similar to parent ratings. Once again, the Sensitive algorithm yielded the greatest sensitivity (.77) and NPV (.74), and lowest specificity (.60) and PPV (.64), whereas the Observed algorithm produced the greatest specificity (.98) and PPV (.89), but the lowest sensitivity (.21) and NPV (.58). The Classic algorithm yielded classification accuracy values between the aforementioned algorithms (see Table 3-5), with sensitivity and specificity estimates of .44 and .88, respectively, and PPV and NPV values of .77 and .63, respectively.

**Empirically-Derived ADHD Algorithms: Concurrent Associations with Functional Impairment**

After determining the optimal threshold for inattention and hyperactivity symptoms based on their prediction of ADHD caseness, separately for parents and teachers, we performed separate multiple regression analyses to compare the Classic, Sensitive, and Observed algorithms
versus DSM-IV diagnoses of ADHD derived from the DISC-IV in predictions of functional impairment and psychopathology from the CBCL and TRF. To combat shared source variance, CBCL and TRF T-scores were averaged to create composite measures of academic functioning (CBCL School and TRF Academic subscales), Social Problems, Internalizing Problems, Externalizing Problems, and Total Problems.

**Academic.** The teacher Sensitive algorithm explained the greatest variance in academic functioning ($\beta=-.38, p<.007, R^2=.17$). Three other algorithms also significantly predicted academic functioning in the following order of variance explained: parent Observed ($\beta=-.36, p<.007, R^2=.14$); parent Sensitive ($\beta=-.33, p<.007, R^2=.11$); and DISC-IV ($\beta=-.32, p<.007, R^2=.11$). The remaining three algorithms did not significantly predict academic functioning: teacher Classic ($\beta=-.27, p<.05, R^2=.09$); parent Classic ($\beta=-.25, p<.05, R^2=.08$); and teacher Observed ($\beta=-.24, p<.05, R^2=.08$).

**Social Problems.** The parent Sensitive algorithm explained the greatest variance in social problems ($\beta=.39, p<.007, R^2=.22$). Five other symptom algorithms also significantly predicted social problems in the following order of variance explained: DISC-IV ($\beta=.36, p<.007, R^2=.21$); teacher Sensitive ($\beta=.34, p<.007, R^2=.20$); parent Observed ($\beta=.34, p<.007, R^2=.20$); parent Classic ($\beta=.29, p<.007, R^2=.17$); and teacher Classic ($\beta=.29, p<.007, R^2=.16$). The teacher Observed algorithm, on the other hand, did not significantly predict social problems ($\beta=.10, p>.05, R^2=.10$).

**Internalizing Problems.** The teacher Classic algorithm was most strongly associated with internalizing problems ($\beta=.37, p<.007, R^2=.19$). The other symptom algorithms that also significantly predicted internalizing problems in the following order of variance explained were: teacher Sensitive ($\beta=.36, p<.007, R^2=.17$); parent Sensitive ($\beta=.31, p<.007, R^2=.12$); and parent
Observed ($\beta=.30, p<.007, R^2=.12$). The other three algorithms did not significantly predict internalizing problems: parent Classic ($\beta=.26, p<.05, R^2=.10$); DISC-IV ($\beta=.21, p>.05, R^2=.07$); and teacher Observed ($\beta=.12, p>.05, R^2=.06$).

**Externalizing Problems.** The teacher Sensitive algorithm explained the greatest variance in externalizing problems ($\beta=.37, p<.007, R^2=.44$). Three other algorithms also significantly predicted externalizing problems in the following order of variance explained: teacher Sensitive ($\beta=.29, p<.007, R^2=.39$); teacher Classic ($\beta=.28, p<.007, R^2=.38$); and teacher Observed ($\beta=.24, p<.007, R^2=.38$). However, the parent Observed ($\beta=.20, p<.05, R^2=.36$); DISC-IV ($\beta=.22, p<.05, R^2=.35$; and parent Classic ($\beta=.17, p>.05, R^2=.35$) algorithms did not significantly predict externalizing problems.

**Total Problems.** The teacher Sensitive ($\beta=.55, p<.007, R^2=.46$) accounted for the greatest variance in total problems. All other algorithms also significantly predicted total problems in the following order of variance explained: parent Sensitive ($\beta=.58 p<.007, R^2=.45$) teacher Classic ($\beta=.47, p<.007, R^2=.37$); DISC-IV ($\beta=.47, p<.007, R^2=.36$); parent Classic ($\beta=.45 p<.007, R^2=.36$); parent Observed ($\beta=.42, p<.007, R^2=.34$); and teacher Observed ($\beta=.28, p<.007, R^2=.25$).

**Discussion**

Given that improving diagnostic procedures for ADHD based on multi-informant ratings will facilitate intervention and prevention efforts, we empirically evaluated the psychometric properties of parent- and teacher-rated ADHD symptoms in an ethnically diverse sample of 7- to 12-year-old children. We identified the optimal threshold for each individual inattention and hyperactivity symptom (separately for parents and teachers) based on its prediction of ADHD diagnostic status. We compared the classification accuracy of three separate symptom algorithms
for parents and teachers: Observed (based on psychometrics), Classic (symptom endorsed as “pretty much” or “very much”), and Sensitive (symptoms endorsed as “just a little,” “pretty much” or “very much”). Finally, we assessed how well the Observed, Classic, and Sensitive algorithms, relative to the DISC-IV, predicted independent measures of academic and social functioning, in addition to internalizing, externalizing, and total problems scales from the CBCL and TRF. Whereas the Sensitive algorithm (teacher) explained the greatest variance in academic functioning, externalizing, and total problems, the Sensitive (parent) and Classic (teacher) models best predicted social problems and internalizing problems, respectively.

Overall, these results converge with previous evidence that individual ADHD symptoms differentially predict ADHD diagnostic status, and that non-DSM-IV algorithms may be superior to DSM-IV in ruling in or ruling out ADHD. Specifically, whereas the Sensitive method for using parent and teacher symptom ratings yielded the greatest sensitivity and NPV, the Observed method for using parent and teacher symptom ratings, which set the strictest symptom thresholds, provided the greatest specificity and PPV in predicting ADHD diagnostic status according to the DISC. Interestingly, the Classic symptom algorithm effectively balanced (i.e., psychometric estimates fell between the Classic and Sensitive methods) sensitivity and specificity, as well as PPV and NPV for both informants. The lower classification accuracy values associated with individual teacher vs. parent ratings of ADHD may reflect the fact that ADHD caseness was based exclusively on the parent-completed DISC-IV (e.g., Achenbach, 1987). Therefore, the classification accuracy of each algorithm should be interpreted within informants (e.g. comparing parent algorithms with each other, and not with teacher algorithms). Finally, the three symptom algorithms differentially predicted functional impairment and psychopathology: the two most sensitive models (Classic and Sensitive) contributed the best fitting models for
predicting multi-domain functioning and psychopathology. Additionally, we note that teacher algorithms were part of the best fitting models for all but one outcome. This may be due in part to teachers having many more opportunities to observe peers in academic and social settings (Gresham, 1982), and therefore may develop greater expertise in identifying problem behaviors. Moreover, given that we included parent-rated ODD symptoms as a covariate, the teacher-rated ADHD symptom algorithms may have provided greater incremental validity relative to parent-rated ADHD symptom algorithms due to potential shared method variance with ODD.

The association of alternative symptom algorithms with functional impairment and broad psychopathology suggests that some ADHD symptoms may better predict ADHD caseness at different thresholds. This formulation is consistent with IRT studies of school-aged (Gomez, Vance, and Gomez, 2009; Gomez, 2008) and preschool children (Purpura et al., 2010) where some ADHD symptoms better discriminated the latent ADHD trait. Similar psychometric differences were reported in a study of 3,402 children and adolescents assessed with the Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Children- Present and Lifetime Version (Kaufman et al., 2000) where some symptoms reflected higher levels of depression and were more discriminating than others (Cole et al., 2011). Additionally, applying IRT-based information about symptom differences in this same study reduced the standard error and increased inter-rater reliability relative to the DSM approach. Thus, research must continue to psychometrically evaluate ADHD symptoms to improve diagnostic accuracy. Algorithms that reflected greater sensitivity for predicting ADHD status generally yielded the best fitting models in the current study. However, we note that the case-control design likely contributed to this phenomenon given that half the participants met full ADHD diagnostic criteria, more lenient thresholds usefully predicted ADHD caseness. Although the Sensitive method may accurately
identify children with ADHD in clinical settings, other algorithms (e.g., Classic) may be superior in other contexts (e.g., schools). Future studies should continue to refine assessment procedures in multiple settings to evaluate the utility of alternative algorithms to improve context-specific diagnostic accuracy.

The current study uniquely compared empirically-derived symptom algorithms for ADHD and compared them to ADHD derived from the DISC-IV with respect to predictions of functional impairment. Future research must further interrogate these (and other) alternative algorithms against DSM-based designations in predictions of socio-emotional, academic, and related areas of functioning. Unlike available standards to evaluate evidence-based interventions (Chambless & Ollendick, 2001), there is less consensus around best practices in psychological assessment. Because ADHD symptoms can be combined in thousands of different combinations to meet diagnostic criteria, and given that ADHD symptoms are only modestly associated with impairment (Fabiano et al., 2006; Pelham et al., 2005), the validity of alternative symptom algorithms should transcend positive and negative predictions of ADHD caseness. Thus, we tested the ability of each algorithm to explain variance in academic, socio-emotional, and behavioral domains. With additional validation, identifying and implementing these symptom algorithms may improve treatment planning by allowing clinicians to target the most clinically significant symptoms (i.e., symptoms most strongly related to academic and social impairment). This is particularly important given that families seeking treatment for ADHD are often more concerned about impairment than symptoms per se (Pelham et al., 2005).

We note several important limitations of this study. First, because these associations were cross-sectional, innovative assessment strategies must demonstrate predictive validity. Second, the use of parent and teacher data to ascertain ADHD and impairment criteria may have inflated
associations through shared method and source variance (Campbell & Fiske, 1959; Gomez et al., 2003), although this motivated the decision to combine parent and teacher ratings of functional impairment. Cross-validating symptom algorithms developed in the current study with other impairment measures (e.g., peer nominations of functioning) will be an important future direction. Similarly, the use of community or nationally representative samples, which provide better estimates of symptom base rates, will increase their generalizability of psychological assessment outside clinic-referred contexts.

The current study evaluated the association of three ADHD symptom algorithms (separately for parents and teachers; i.e., six total) with diverse measures of functional impairment and broad psychopathology relative to ADHD derived from the DISC-IV conducted with the parent. The two most sensitive algorithms (i.e., Classic and Sensitive) based on parent and teacher ratings (i.e., four algorithms) better predicted academic and social functioning, as well as internalizing, externalizing, and total problems relative to DISC-IV diagnoses. Future research should examine symptom utilities to determine how individual items should be optimally weighted and combined, whether certain symptom algorithms are most effective in different contexts, as well as develop alternative ADHD algorithms and test them against DISC-IV diagnoses with regard to predicting functioning and prospective outcomes.
<table>
<thead>
<tr>
<th>Variable</th>
<th>ADHD</th>
<th>Comparison</th>
<th>t/χ²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 42)</td>
<td>(n = 42)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>9.52 (1.27)</td>
<td>9.49 (1.32)</td>
<td>.13</td>
</tr>
<tr>
<td>Number of Boys (%)</td>
<td>35 (83%)</td>
<td>28 (67%)</td>
<td>2.9</td>
</tr>
<tr>
<td>Number of non-White children (%)</td>
<td>23 (55%)</td>
<td>18 (49%)</td>
<td>.29</td>
</tr>
<tr>
<td>Number of DISC-IV ADHD Symptoms</td>
<td>12.55 (3.37)</td>
<td>2.74 (2.72)</td>
<td>14.79*</td>
</tr>
<tr>
<td>Number of DISC-IV ODD Symptoms</td>
<td>2.26 (2.13)</td>
<td>1.10 (1.59)</td>
<td>2.84*</td>
</tr>
<tr>
<td>Number of DISC-IV CD Symptoms</td>
<td>.43 (.80)</td>
<td>.24 (.48)</td>
<td>1.32</td>
</tr>
<tr>
<td>Academic Functioning †</td>
<td>43.39 (8.87)</td>
<td>50.05 (8.32)</td>
<td>-3.48*</td>
</tr>
<tr>
<td>Social Problems †</td>
<td>59.68 (6.35)</td>
<td>55.06 (5.38)</td>
<td>3.62*</td>
</tr>
<tr>
<td>Internalizing Problems †</td>
<td>55.60 (8.72)</td>
<td>52.72 (7.71)</td>
<td>1.61</td>
</tr>
<tr>
<td>Externalizing Problems †</td>
<td>54.80 (7.39)</td>
<td>51.05 (6.79)</td>
<td>2.44*</td>
</tr>
<tr>
<td>Total Problems †</td>
<td>60.07 (6.41)</td>
<td>52.31 (7.55)</td>
<td>5.10*</td>
</tr>
</tbody>
</table>

*Note.* DISC-IV = Diagnostic Interview Schedule for Children, 4th edition; ADHD = attention-deficit/hyperactivity disorder; ODD = oppositional defiant disorder; CD = conduct disorder; *p < .05.

*a* Ethnicity data missing for 5 control children.

†Average of CBCL and TRF subscales’ T-scores.
Table 3-2. Total predictive power values (TPVs) for parent-rated ADHD symptoms

<table>
<thead>
<tr>
<th>Inattention Items</th>
<th>TPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. is often easily distracted by extraneous stimuli&lt;sup&gt;3&lt;/sup&gt;</td>
<td>.82</td>
</tr>
<tr>
<td>44. is often forgetful in daily activities&lt;sup&gt;2&lt;/sup&gt;</td>
<td>.79</td>
</tr>
<tr>
<td>18. often does not seem to listen when spoken to directly&lt;sup&gt;2&lt;/sup&gt;</td>
<td>.78</td>
</tr>
<tr>
<td>29. often has difficulty sustaining attention in tasks or play activities&lt;sup&gt;2&lt;/sup&gt;</td>
<td>.78</td>
</tr>
<tr>
<td>34. often loses things necessary for tasks or activities&lt;sup&gt;2&lt;/sup&gt;</td>
<td>.78</td>
</tr>
<tr>
<td>23. often fails to give close attention to details or makes careless mistakes in schoolwork, work, or other activities&lt;sup&gt;3&lt;/sup&gt;</td>
<td>.77</td>
</tr>
<tr>
<td>42. often has difficulty organizing tasks and activities&lt;sup&gt;3&lt;/sup&gt;</td>
<td>.76</td>
</tr>
<tr>
<td>27. often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace&lt;sup&gt;2&lt;/sup&gt;</td>
<td>.75</td>
</tr>
<tr>
<td>37. often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort&lt;sup&gt;2&lt;/sup&gt;</td>
<td>.75</td>
</tr>
<tr>
<td>Hyperactivity/Impulsivity Items</td>
<td></td>
</tr>
<tr>
<td>33. is often “on the go” or often acts as if “driven by a motor”&lt;sup&gt;2&lt;/sup&gt;</td>
<td>.82</td>
</tr>
<tr>
<td>12. often fidgets with hands or feet or squirms in seat&lt;sup&gt;3&lt;/sup&gt;</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>22. often has difficulty playing or engaging in leisure activities quietly</td>
<td>.80</td>
</tr>
<tr>
<td>25. often has difficulty playing or engaging in leisure activities quietly</td>
<td>.80</td>
</tr>
<tr>
<td>19. often blurts out answers before questions have been completed</td>
<td>.79</td>
</tr>
<tr>
<td>30. often has difficulty awaiting turn</td>
<td>.79</td>
</tr>
<tr>
<td>35. often runs about or climbs excessively in situations in which it is inappropriate</td>
<td>.78</td>
</tr>
<tr>
<td>7. often talks excessively</td>
<td>.77</td>
</tr>
<tr>
<td>1. often interrupts or intrudes on others</td>
<td>.76</td>
</tr>
</tbody>
</table>

Note. ADHD = attention-deficit/hyperactivity disorder

1 A rating of “just a little,” “pretty much,” or “very much” was most predictive of ADHD and considered a symptom endorsement.

2 A rating of “pretty much” or “very much” was most predictive of ADHD and considered a symptom endorsement.

3 A rating of “very much” was most predictive of ADHD and considered a symptom endorsement.
Table 3-3. Total predictive power values (TPVs) for teacher-rated ADHD symptoms

<table>
<thead>
<tr>
<th>Inattention Items</th>
<th>TPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>29. often has difficulty sustaining attention in tasks or play activities (^2)</td>
<td>.78</td>
</tr>
<tr>
<td>18. often does not seem to listen when spoken to directly (^2)</td>
<td>.75</td>
</tr>
<tr>
<td>9. is often easily distracted by extraneous stimuli (^3)</td>
<td>.74</td>
</tr>
<tr>
<td>27. often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (^2)</td>
<td>.72</td>
</tr>
<tr>
<td>37. often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (^2)</td>
<td>.72</td>
</tr>
<tr>
<td>42. often has difficulty organizing tasks and activities (^2)</td>
<td>.72</td>
</tr>
<tr>
<td>34. often loses things necessary for tasks or activities (^3)</td>
<td>.71</td>
</tr>
<tr>
<td>44. is often forgetful in daily activities (^3)</td>
<td>.71</td>
</tr>
<tr>
<td>23. often fails to give close attention to details or makes careless mistakes in schoolwork, work, or other activities (^3)</td>
<td>.70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hyperactivity/Impulsivity Items</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12. often fidgets with hands or feet or squirms in seat (^2)</td>
<td>.75</td>
</tr>
<tr>
<td>30. often has difficulty awaiting turn (^2)</td>
<td>.74</td>
</tr>
</tbody>
</table>
35. often runs about or climbs excessively in situations in which it is inappropriate\(^1\) .74

25. often leaves seat in classroom or in other situations in which remaining seated is expected\(^3\) .73

33. is often "on the go" or often acts as if "driven by a motor"\(^2\) .73

7. often talks excessively\(^3\) .72

19. often blurts out answers before questions have been completed\(^3\) .71

1. often interrupts or intrudes on others\(^2\) .70

22. often has difficulty playing or engaging in leisure activities quietly\(^2\) .70

Note. ADHD = attention-deficit/hyperactivity disorder

\(^1\)A rating of “just a little,” “pretty much,” or “very much” was most predictive of ADHD and considered a symptom endorsement.

\(^2\)A rating of “pretty much” or “very much” was most predictive of ADHD and considered a symptom endorsement.

\(^3\)A rating of “very much” was most predictive of ADHD and considered a symptom endorsement.
### Table 3-4. Number of individuals meeting ADHD symptom criteria according to each symptom algorithm

<table>
<thead>
<tr>
<th>Symptom Algorithm</th>
<th>Parent</th>
<th>Teacher&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Classic</td>
<td>Observed</td>
</tr>
<tr>
<td>ADHD (n = 42)</td>
<td>Present</td>
<td>Absent</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>Comparison (n = 42)</td>
<td>2</td>
<td>41</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>53</td>
</tr>
</tbody>
</table>

<sup>a</sup>Teacher-rated Disruptive Behavior Disorder Rating Scale data available for 82 children.

*Note. ADHD = attention-deficit/hyperactivity disorder; Parent = parent-rated ADHD symptoms; Teacher = teacher-rated ADHD symptoms; Classic = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “pretty much” or “very much”; Observed = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed at their optimal (i.e., highest total predictive value) level; Sensitive = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “just a little,” “pretty much” or “very much”.*
Table 3-5. Classification accuracy of each ADHD symptom algorithm\textsuperscript{a}

<table>
<thead>
<tr>
<th>Symptom Algorithm</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parent\textsuperscript{b}</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classic</td>
<td>.71</td>
<td>.95</td>
<td>.94</td>
<td>.77</td>
</tr>
<tr>
<td>Observed</td>
<td>.46</td>
<td>.98</td>
<td>.95</td>
<td>.66</td>
</tr>
<tr>
<td>Sensitive</td>
<td>1.00</td>
<td>.59</td>
<td>.68</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Teacher\textsuperscript{b}</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classic</td>
<td>.44</td>
<td>.88</td>
<td>.77</td>
<td>.63</td>
</tr>
<tr>
<td>Observed</td>
<td>.21</td>
<td>.98</td>
<td>.89</td>
<td>.58</td>
</tr>
<tr>
<td>Sensitive</td>
<td>.77</td>
<td>.60</td>
<td>.64</td>
<td>.74</td>
</tr>
</tbody>
</table>

\textit{Note}. ADHD = attention-deficit/hyperactivity disorder; PPV = positive predictive value; NPV = negative predictive value; Parent = parent-rated ADHD symptoms; Teacher = teacher-rated ADHD symptoms; Classic = \( \geq 6 \) ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “pretty much” or “very much”; Observed = \( \geq 6 \) ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed at their optimal (i.e., highest total predictive value) level; Sensitive = \( \geq 6 \) ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “just a little,” “pretty much” or “very much”.

\textsuperscript{a}Classification accuracy based on DISC-IV ADHD diagnoses.

\textsuperscript{b}Symptoms rated on the Disruptive Behavior Disorder Rating Scale.
<table>
<thead>
<tr>
<th>Symptom Algorithm</th>
<th>Academic Functioning&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Social Problems&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Internalizing Problems&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Externalizing Problems&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Total Problems&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DISC-IV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.32**</td>
<td>.11</td>
<td>.36**</td>
<td>.21</td>
<td>.21</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Parent</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.25*</td>
<td>.08</td>
<td>.29**</td>
<td>.17</td>
<td>.26*</td>
<td>.10</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.36**</td>
<td>.14</td>
<td>.34**</td>
<td>.20</td>
<td>.30**</td>
<td>.12</td>
</tr>
<tr>
<td>Sensitive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.33**</td>
<td>.11</td>
<td>.39**</td>
<td>.22†</td>
<td>.31**</td>
<td>.12</td>
</tr>
<tr>
<td><strong>Teacher</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.27*</td>
<td>.09</td>
<td>.29**</td>
<td>.16</td>
<td>.37**</td>
<td>.19†</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.24*</td>
<td>.08</td>
<td>.10</td>
<td>.10</td>
<td>.12</td>
<td>.06</td>
</tr>
<tr>
<td>Sensitive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.38**</td>
<td>.17†</td>
<td>.34**</td>
<td>.20</td>
<td>.36**</td>
<td>.17</td>
</tr>
</tbody>
</table>

*β* indicates the beta coefficient, *R<sup>2</sup>* indicates the coefficient of determination.
Note. ADHD = attention-deficit/hyperactivity disorder; DISC-IV = Diagnostic Interview Schedule for Children, 4th edition; Parent = parent-rated ADHD symptoms; Teacher = teacher-rated ADHD symptoms; Classic = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “pretty much” or “very much”; Observed = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed at their optimal (i.e., highest total predictive value) level; Sensitive = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “just a little,” “pretty much” or “very much.”

*p < .05.

**p ≤ .007 (α value based on Bonferroni correction).

Average of CBCL and TRF subscales’ T-scores.

†Highest R² value within a functional domain.
CHAPTER FOUR: CROSS-VALIDATION AND DEVELOPMENT OF EMPIRICALLY-DERIVED ADHD ASSESSMENT STRATEGIES: INSIGHTS FROM THE NATIONAL LONGITUDINAL STUDY OF ADOLESCENT HEALTH (ADD HEALTH)

There is replicated evidence that individual ADHD symptoms differentially predict ADHD diagnostic status, and that non DSM-based symptom algorithms are superior to DSM criteria for ruling in or ruling out an ADHD (Lahey et al., 1994; Power et al., 2001; Shemmassian & Lee, 2014). Perhaps because all 18 ADHD symptoms vary with respect to reflecting individual differences in the latent ADHD trait, certain items may better predict ADHD caseness. This formulation is consistent with item response theory (IRT) studies of ADHD symptom ratings in school-age (Gomez et al., 2009; Gomez, 2008) and preschool children (Purpura et al., 2010). Thus, prosecuting the utility of symptoms using novel approaches may not only identify relatively more and less useful symptoms, but also improve diagnostic accuracy. Identifying the most predictive symptoms, at their optimal severity levels (i.e., greatest predictive utility) is especially important given that some ADHD symptoms lack social validity (i.e., poor descriptors of treatment outcomes; Pelham & Fabiano, 2001). Another potential implication is that the accurate determination of ADHD caseness may improve if specific symptoms were required rather than the DSM approach, which simply requires any 6 (or more) symptoms. Therefore, previous investigations (e.g., Lahey et al., 1994; Power et al., 2001; Shemmassian & Lee, 2014) found that particular symptoms, at certain thresholds, were superior for ruling in or ruling out ADHD; however, this is not reflected in DSM-IV. Moreover, the new DSM-5 diagnostic criteria for ADHD did not address these issues, leaving symptom criteria essentially unchanged. Revising assessment procedures to reflect greater weight for psychometrically superior symptoms, or requiring endorsement specific items and/or at specific
severity levels, may improve diagnostic accuracy, reduce diagnostic heterogeneity, and make diagnostic procedures more efficient.

To our knowledge, outside of the proposed study in Chapter 3, no previous study has directly compared an ADHD algorithm constructed according to psychometric properties of symptoms to the DSM method (or similar structured diagnostic interview) in predicting functional impairment. Although alternative algorithms must empirically outperform DSM-based designations in predictions of socio-emotional, academic, and related areas of functioning, there are no systematic standards for judging best practices in psychological assessment, unlike the standards used to evaluate evidence-based interventions (Chambless & Ollendick, 2001). Because ADHD symptoms can be combined in thousands of different combinations to meet diagnostic criteria, and given that ADHD symptoms are only modestly associated with impairment (Fabiano et al., 2006; Pelham et al., 2005), evaluating the validity of alternative symptom algorithms should include, but not be limited to, positive and negative predictions of ADHD caseness as well as significant explanatory power of multi-domain functional impairment. Identifying and implementing these symptom algorithms may improve treatment planning by allowing clinicians to target the most significant behaviors.

In addition to comparing empirically-derived assessment strategies against DSM-IV for predicting ADHD-related functional impairment, it is important that any preliminary algorithm be cross-validated. That is, replicating findings in independent and representative samples will increase the generalizability of using specific symptom algorithms identified in one study as having the greatest predictive utility and explaining the most variance in functional outcomes. Generalizability of these methods will address a number of important limitations common to clinical research. First, many ADHD studies probe parents and teachers to ascertain diagnostic
status and impairment across multiple domains (e.g., Lee et al., 2008; Shemmassian & Lee, 2012). This method may inflate associations between symptomatology and impairment because of shared method and source variance (Campbell & Fiske, 1959; Gomez et al., 2003). Cross-validating symptom algorithms minimizes this potential inflation because the parent and teacher ratings used to develop the algorithms in one sample are conservatively tested in predictions of impairment that are rated by different informants. Second, given that alternative ADHD assessment strategies often depend on endorsing specific symptoms at certain severity thresholds, they suffer from fairly low base rates when they require several symptoms to be endorsed at higher severity levels. Although the use of kappa statistics for PPP and NPP mitigates some variations in predictability across settings with different base rates (Frick et al., 1994), explicit evaluation of alternative strategies in different samples remains critical.

A third important question with respect to cross-validation of empirically-derived ADHD assessment strategies is how ADHD symptoms perform across different sampling strategies (e.g., clinical vs. epidemiological). Given that the validation of ADHD assessment strategies has been conducted predominantly with clinic-referred children, the prevalence and utility of symptoms and symptom algorithms must be appraised in non-clinical settings and across development. Optimal algorithms developed in multiple contexts (e.g., clinic-referred settings, population-based samples) may reflect the inherent differences in these populations (e.g., oversampling of probands in clinical samples) and may be more accurate within those settings. Although ADHD is primarily characterized as a childhood disorder, recent research suggests a latent ADHD trait that may be exhibited through different behaviors or at different levels across development and across settings (Normand, Flora, Toplak, & Tannock, 2012). Therefore, ADHD symptoms may differentially perform across development and contexts. For example, in an IRT study of parent-
and teacher-rated ADHD symptoms in a large sample of school-aged children, the overall reliability of the DSM-IV ADHD rating scale used was poor for latent scores below the mean (i.e., -1 SD) and well-above the mean (i.e., +3 SD) (Gomez, 2008). IRT studies of ADHD symptoms have also revealed poor reliability estimates at high and low latent trait scores in samples of pre-school-aged children (Purpura et al., 2010) and adults (Gomez, 2011), suggesting that ADHD rating scales may have poor validity for high-risk and typically developing populations. However, we know of no studies of the prevalence of specific symptoms or that evaluated the differential item functioning of symptom ratings across multiple age groups, despite the fact that ADHD symptoms are highly sensitive to development (i.e., persist or desist) (Wilens, Faraone, & Biederman, 2004). A meta-analysis of 32 follow-up studies of ADHD reported that the prevalence rate of ADHD in young adulthood (age 25) largely depended on the definition of persistence (Faraone, Biederman, & Mick, 2006). When ADHD was defined according to full diagnostic criteria, persistence rates were 15% whereas studies that included cases of ADHD in partial remission (i.e., some impairing symptoms but below formal diagnostic criteria), persistence approached 65%. However, despite the well-documented persistence of ADHD symptoms, (e.g., Faraone & Biederman, 2005; Kooji et al., 2005) and ADHD-related impairment into adulthood (Barkley et al., 2008; Lee et al., 2011), knowledge about the persistence of specific symptoms is less clear. Moreover, although many symptoms decline in severity or remit with age (Wilens et al., 2004), certain symptoms may be more or less prevalent throughout the lifespan, influencing the presentation of the disorder across development.

Population-based samples facilitate estimates of the prevalence and psychometric properties of ADHD symptoms across development and other relevant contexts (e.g., SES, race-ethnicity). Moreover, evaluating whether symptom algorithms developed with case-control or
clinic-referred studies accurately identify functionally-impaired youth in an unselected sample is important. To date, studies in such samples have been limited to factor analytic work and the development of normative data for parent- (DuPaul et al., 1998) and teacher-rated symptoms (Du Paul et al., 1997). Thus, there is a need to extend empirically-derived symptom algorithms beyond case-control and over-sampled studies to evaluate their validity for use with more representative populations. Successful validation will improve the efficiency and accuracy of ADHD assessment across these groups. The goal of this study was to evaluate the predictive utility of empirically-derived ADHD symptom algorithms based on a case-control study of parent- and teacher-ratings of school-aged children (Sample 1; see Chapter 3) in a population-based sample (Sample 2). We also developed ADHD symptom algorithms separately in the population-based sample and compared them to the DSM-IV approach for identifying adolescents experiencing multi-domain functional impairment (Specific Aim 3).

Method

Participants

Data are from the National Longitudinal Study of Adolescent Health (Add Health; http://www.cpc.unc.edu/projects/addhealth), a longitudinal, population-based study of grade 7-12 adolescents in the United States. A systematic random sample of 80 high schools was selected from all high schools in the United States that had an 11th grade and at least 30 enrollees. Schools were selected based on enrollment size, stratified by region, urbanicity, school type, and percentage white. Each high school’s largest feeder (mostly middle schools) was also recruited. Information about participants’ functioning was gathered via in-home interviews during Wave 1. Analyses were performed in Stata using the Grand Sample Weight at Wave 1. Additional details about the Wave 1 study design, and in-home sample and interview are provided in Resnick et al.,
ADHD symptoms were self-reported retrospectively during Wave 3 of the study, from 2001-2002 (http://www.cpc.unc.edu/projects/addhealth/design/wave3) when participants were 18-28 years old. Participants in our sample consisted of 11,247 (46% male) ethnically diverse 10- to 19-year-old individuals (M = 14.98, SD = 1.70; 98.8% of the sample was 12-18 years) at Wave 1. 55.5% of participants self-identified as White, 22.9% as Black/African American, 5.5% as American Indian or Native American, 9.5% as Asian or Pacific Islander, and 8.9% as Other. Moreover, 18.1% of individuals self-identified as Hispanic or Latino (n = 10,659). 827 participants met symptom criteria for ADHD compared to 10,420 typically-developing participants. Although these two groups were comparable on age, expectedly, youth who met symptom criteria for ADHD were significantly more likely to be male, White, have lower great point averages (GPA), experience greater health impairment, and reported less parental closeness (further description is provided below and in Table 4-1).

**Measures**

Although the total sample consisted of 11,247 participants, complete data varied by domain. The number of individuals who provided valid and usable outcome data for each domain is described for each measure.

**ADHD.** At Wave 3, participants (ages 18-26) were asked to retrospectively rate the severity of their ADHD symptoms from 5-12 years (9 inattention and 8 hyperactivity symptoms; 1 DSM-IV impulsivity symptom, “often interrupts or intrudes on others,” was not queried). Items were rated as “never or rarely,” “sometimes,” “often” or “very often” using an in-home interview developed exclusively for Add Health. Symptoms were considered present when rated as “often” or “very often” (Lahey et al., 1998; Pelham et al., 1992) and Cronbach’s alphas were .80 and .74 for the number of inattention and hyperactivity symptoms, respectively. Participants were placed
in the ADHD group (n = 827) if they met symptom criteria for any subtype of ADHD. 32% (n = 265) met symptom criteria for the Predominantly Inattentive Type, 39% (n = 324) for the Predominantly Hyperactive/Impulsive Type, and 29% (n = 238) for the Combined Type.

**GPA** (n = 10,179; 91% available data). Participants self-reported their grades (“A,” “B,” “C,” or “D or lower,” coded as “1,” “2,” “3,” and “4,” respectively) in English/Language Arts, Mathematics, History/Social Studies, and Science were during the most recent grading period. Participants’ grades were reverse scored to yield traditional grade point data (i.e., (“A,” “B,” “C,” or “D or lower,” were coded as “4,” “3,” “2,” and “1,” respectively). Participants’ grades were averaged across subjects to determine their GPA. Participants with missing grade information for two or more subjects were excluded for this set of analyses.

**Health Interference** (n = 10,909; 97% available data). Participants endorsed the frequency of a health or emotional problem that caused them to miss a day of school in the last month (“never,” “just a few times,” “about once a week,” “almost every day,” and “every day”); and how often a health or emotional problem caused them to miss a social or recreational activity in the last month (“never,” “just a few times,” “about once a week,” “almost every day,” and “every day”). These two questions were combined to form a composite measure of how often a health or emotional problem interfered with participants’ attendance in any activity. These two questions yielded a Cronbach’s alpha value of .54.

**Parental Closeness** (n = 10,788; 96% available data). Participants separately rated how close they felt with their mothers and fathers, and separately whether their parents cared about them (“never or rarely,” “very little,” “somewhat,” “quite a bit,” or “very often”). Participants’ responses were averaged across both questions and for both parents to form a single composite measure of parental closeness. The Cronbach’s alpha was .72 for these four items.
Data Analyses

Symptom Severity Thresholds

Although individual ADHD symptoms differentially predict ADHD diagnostic status at different levels for parents and teachers (Pelham et al., 2005; Shemmassian & Lee, 2014), there is a particular lack of knowledge about the prevalence or predictive utility of specific ADHD symptoms based on self-report, or in population-based samples. Therefore, we began by calculating the base rate of each ADHD symptom when endorsed at the “often” or “very often” levels (Table 4-2). We then calculated total predictive value (TPV; i.e., hit rate) of each level (i.e., “never or rarely,” “sometimes,” “often,” “very often”) of each ADHD symptom against ADHD versus non-ADHD status (see Measures section, above). For example, we calculated TPVs associated with three comparisons: (1) “never or rarely” versus “sometimes,” “often,” or “very often;” (2) “never or rarely” or “sometimes” versus “often” or “very often;” and (3) “never or rarely,” “sometimes,” and “often” versus “very often.” The symptom level at which each item yielded the highest TPV for ADHD caseness was considered optimal (Table 4-3).

Symptom Algorithms and Classification Accuracy

Next, we created five separate ADHD algorithms: (1) Classic: ≥ 6 of 9 ADHD inattention and/or hyperactivity symptoms endorsed as “often” or “very often” (i.e., akin to the DSM-IV method; Lahey et al., 1998; Pelham et al., 1992); (2) Observed: ≥ 6 of 9 inattention and/or hyperactivity symptoms endorsed at their highest TPV level; (3) Sensitive: ≥ 6 of 9 inattention and/or hyperactivity symptoms endorsed as “sometimes,” “often” or “very often;” (4) Specific: ≥ 6 of 9 inattention and/or hyperactivity symptoms endorsed as “very often”; and (5) Clinical: ≥ 6 of 9 inattention and/or hyperactivity symptoms endorsed at the highest TPV level obtained from parent symptom ratings in the case-control study from Chapter 3 (see Table 3-2). The highest
TPV for 82% of symptoms in the current study was evident when symptoms were endorsed at the “very often” level (see Table 4-3). The numbers of individuals that did or did not meet symptom criteria according to each algorithm are presented in Table 4-4. We also calculated the sensitivity, specificity, PPV, and NPV of each symptom algorithm based on their correspondence with ADHD diagnostic status according to the Classic method (Table 4-5).

**Associations with Functional Impairment**

Although individual ADHD symptoms differentially predict ADHD at different levels for parents and teachers (Pelham et al., 2005), it is unclear whether different combinations of symptom levels and informants are superior to the Classic method in predictions of functional impairment. Moreover, we know of no study where the utility of empirically-derived symptom algorithms developed from a selected sample (e.g., case-control) were then separately evaluated in a population-based sample. The final evaluation consisted of separately testing the validity of the five symptom algorithms with three functional outcomes (i.e., academic, familial, and general health). Because ADHD was not diagnostically determined in this sample, these functional outcomes are unrelated to ADHD symptom criteria status. Specifically, we performed multiple regressions separately for each symptom algorithm controlling for age and sex at Wave 1 (i.e., 15 regressions total). We then interpreted the $R^2$ values of statistically significant regression coefficients to determine which algorithm explained the greatest amount of variance for each outcome variable (higher values correspond to greater variance explained) (Table 4-6). Bonferroni corrections were applied to each set of regressions (i.e., for each functional domain; $\alpha=.05/5=.01$) to control for multiple comparisons.

**Results**

*Base Rates of ADHD Symptoms*
The base rates of the nine inattention symptoms endorsed as “often” or “very often” ranged from 8% (“When you were between 5 and 12, you avoided, disliked, or were reluctant to engage in work requiring sustained mental effort”) to 23% (“When you were between 5 and 12, you were easily distracted”). The base rates of the eight hyperactivity symptoms endorsed as “often” or “very often” ranged from 8% (“You left your seat in the classroom or in other situations when being seated was expected”) to 32% (“You fidgeted with your hands or feet or squirmed in your seat”) (see Table 4-2 for base rates).

Clinical Utility of Individual ADHD Symptoms

7 of 9 (78%) and 7 of 8 (88%) ADHD inattention and hyperactivity symptoms, respectively, were most predictive of ADHD status when endorsed as “very often.” In addition, two of the inattention symptoms, and one hyperactivity symptom were most predictive of ADHD status when endorsed as “often” or “very often,” (see Table 4-3 for TPV values and cutoffs).

Classification Accuracy of ADHD Symptom Algorithms

Four symptom algorithms (Observed, Sensitive, Specific, Clinical; see Symptom Algorithms and Classification Accuracy section, above) were applied to classify children as meeting symptom criteria for any ADHD subtype according to the Classic Method. The Sensitive algorithm, which relaxed the symptom threshold, yielded the greatest sensitivity (1.00) and NPV (1.00), but the lowest specificity (.54) and PPV (.15). The Observed algorithm, constructed from symptom thresholds based on each symptoms’ predictive utility, most of which were most predictive when endorsed at the “very often” level, yielded maximum specificity (1.00) and PPV (1.00), high NPV (.94), but low sensitivity (.23). The Specific algorithm, which set the most stringent symptom threshold, also yielded the maximum specificity (1.00) and PPV (1.00), high NPV (.94), but the lowest sensitivity (.19). Finally, the Clinical algorithm,
constructed from symptom thresholds based on the predictive utility obtained from parent symptom ratings in the case-control study from Chapter 3, most of which were most predictive when endorsed at the “often” level, yielded the second-highest sensitivity (.74), specificity (.99), PPV (.91), and NPV (.98) (see Table 4.5).

**Empirically-Derived ADHD Algorithms: Associations with Functional Impairment**

After determining the optimal threshold for inattention and hyperactivity symptoms based on their prediction of ADHD status to create the Observed algorithm, we performed separate multiple regression analyses to compare the Classic, Observed, Sensitive, Specific, and Clinical symptom algorithms of ADHD in predictions of GPA, health interference, and parental closeness.

**GPA.** The Sensitive algorithm explained the greatest variance in GPA ($\beta=-.12.18, p<.01, R^2=.05$). The Classic ($\beta=-7.19, p<.01, R^2=.03$) and Clinical ($\beta=-6.20, p<.01, R^2=.03$) also significantly predicted GPA. On the other hand, the Specific ($\beta=-2.62, p<.05, R^2=.02$) and Observed ($\beta=-2.41, p<.05, R^2=.02$) algorithms did not significantly predict GPA.

**Health Interference.** The Sensitive algorithm explained the greatest variance in health interference ($\beta=4.98, p<.01, R^2=.02$), followed by the Classic algorithm ($\beta=3.43, p<.01, R^2=.02$). The remaining symptom algorithms were unrelated to health interference: Observed ($\beta=2.48, p<.05, R^2=.02$); Specific ($\beta=1.92, p<.05, R^2=.02$); and Clinical ($\beta=1.71, p<.05, R^2=.02$).

**Parental Closeness.** The Sensitive algorithm was most strongly associated with parental closeness ($\beta=-6.40, p<.01, R^2=.04$). Two other symptom algorithms also significantly predicted parental in the following order of variance explained: Classic ($\beta=-4.38, p<.01, R^2=.04$) and Clinical ($\beta=-3.95, p<.01, R^2=.04$). However, the Observed ($\beta=-2.09, p<.05, R^2=.03$) and Specific ($\beta=-1.99, p<.05, R^2=.03$) algorithms did not significantly predict parental closeness.

**Discussion**
There is relatively little knowledge regarding the prevalence and psychometric properties of individual ADHD symptoms and symptom algorithms in population-based samples. Given that ADHD symptoms may perform differently across contexts, it is important to examine whether previous findings, based largely on clinic-referred samples and/or case-control studies, generalize to population-based samples. We calculated the base rates and predictive utility of retrospectively recalled ADHD symptoms in a population-based sample of 11,247 youths. Specifically, we estimated the threshold at which each inattention and hyperactivity symptom best predicted ADHD status (i.e., yielded the highest hit rate). Next, we compared the classification accuracies of various symptom algorithms based on their agreement with a Classic algorithm (symptoms endorsed as “often” or “very often”): Observed (organized according to the psychometric properties in the sample), Sensitive (symptoms endorsed as “sometimes,” “often” or “very often”), Specific (symptoms endorsed as “very often”), and Clinical (based on psychometrics from the case-control study in Chapter 3). Finally, we compared the five algorithms in their ability to predict independent measures of academic functioning (GPA), health interference, and parental closeness. Across all three functional domains, the Sensitive algorithm explained the greatest amount of variance.

As expected, individual ADHD symptoms were variably associated with ADHD status. Moreover, whereas most symptoms typically demonstrate highest predictive utility when endorsed at their two highest (out of four) severity levels in a case-control study (see Chapter 3; Power et al., 2001), symptoms in Add Health were typically most predictive of ADHD when endorsed at the highest severity level. The divergence in optimal severity thresholds for ADHD symptoms across contexts potentially reflects the different base rates of symptom endorsement. The DSM-IV Field Trials for ADHD evaluated the psychometric properties of all candidate
symptoms using a clinic-referred sample of 380 4- to 17-year-old youths (Lahey et al., 1994). Perhaps unsurprisingly then, psychometric properties of ADHD symptoms obtained from clinic-referred samples, where a large number of children meet ADHD symptom criteria or are “subthreshold” cases (i.e., elevated but below full symptom criteria), may not generalize to population-based samples. It is therefore important to further characterize non-clinical populations to determine best practices for larger, community-based settings including schools and primary care.

The accuracy of the different classification approaches also expectedly suggested that different symptom algorithms might be superior for ruling in or ruling out ADHD. Until now, however, classification accuracy estimates of various symptom algorithms have not been reported in larger, population-based samples. Whereas the Sensitive method, which relaxed the symptom threshold, yielded maximum (i.e., 1.00) sensitivity and NPV, the Observed and Specific methods yielded maximum specificity and PPV. Interestingly, the Clinical algorithm effectively balanced (i.e., yielded psychometric estimates between the Sensitive, and Observed and Specific models) sensitivity and specificity, as well as PPV and NPV. Moreover, this algorithm yielded specificity, PPV, and NPV values above .90, with its lowest value (sensitivity; .74) well above the other algorithms’ lowest values (see Table 4-5). The Clinical algorithm’s consistently high accuracy reflects its similarity to the Classic algorithm, since most symptoms’ optimal severity thresholds are achieved at the two highest severity levels (see Chapter 3; Power et al., 2001). Although all of the symptom algorithms in the current study yielded variable classification accuracy estimates, the algorithms themselves were variations of a univariate procedure (i.e., a retrospective ADHD measure) due to the limited availability of relevant assessment data for ADHD. An important next step in population-based samples will be
to use multivariate methods, which better approximate clinical settings, including the use of multiple informants (e.g., parents and teachers) and methods (e.g., rating scales, observational data) data for ascertaining ADHD status (Pineda, Puerta, Aguirre, García-Barrera, & Kamphaus, 2007), especially given that ADHD rating scales have demonstrated more source variance than trait variance (Gomez et al., 2003). Because rating scale data from one informant (i.e., self-report in the current study) may provide weak convergent and discriminant validity regardless of the algorithms used, multi-method approaches might improve the validity of ADHD classification (Pineda et al., 2007), including more accurately identifying functionally impaired youth.

While examining classification accuracy estimates is important to determine the relative utility of each algorithm (e.g., to rule in or rule out diagnoses), promising algorithm(s) must identify functionally impaired youth. Variation across the five symptoms algorithms in the current study is suggested by their differential predictions of functioning. Specifically, the Sensitive algorithm explained the greatest amount of variance in GPA, health interference, and parental closeness. Moreover, the Classic algorithm, the second most sensitive method, was the only other to significantly predict functioning across all three functional domains. Moreover, whereas the Clinical algorithm significantly predicted GPA and parental closeness, the Observed and Specific algorithms did not significantly functioning in any domain. Although the Sensitive algorithm explained the greatest variance across the three domains examined in the current study, these results should be interpreted cautiously given its highly discrepant classification accuracy estimates. Specifically, whereas the other four algorithms each classified no more than 7.3% of individuals as meeting symptom criteria for ADHD, the Sensitive algorithm classified 5,591 individuals (53% of the sample) (see Table 4-4). A recent meta-analysis of 86 studies of children and adolescents (n = 163,688) and 11 studies of adults (n = 14,112) reported pooled DSM-IV
ADHD prevalence estimates ranging from 5.9-7.1% in children and adolescents regardless of the informant probed (i.e., parents or teachers), or a best estimate diagnostic procedure. Moreover, self-report measures in young adults yielded prevalence estimates of 5.0% (Willcutt, 2012). Thus, the Sensitive algorithm may overrepresent ADHD cases in population-based samples. Given that the Classic algorithm identified 7.3% of individuals as meeting symptom criteria for ADHD, and also significantly predicted functioning across all three domains, it may be the single most useful model for evaluating ADHD in population-based samples.

We note several important limitations of this study. First, because 18-28 year-old participants retrospectively reported their ADHD symptoms at ages 5 to 12 years, ratings are susceptible to recall bias (Hassan, 2006). One hyperactivity symptom (“often interrupts or intrudes on others”) was also completely excluded. In addition, outcome variables were self-reported at Wave 1 when participants were 10- to 19-years old. Taken together, this data yielded a quasi-longitudinal design. Moreover, the use of self-report data to ascertain ADHD and functioning criteria may have inflated associated through shared method and source variance (Campbell & Fiske, 1959; Gomez et al., 2003). Nevertheless, to our knowledge, these population-based data provided an innovative examination of the base rates and predictive utility of ADHD symptoms in a non-referred population. Thus, the current study is an important first step for examining the generalizability of findings based mostly on potentially biased clinic-referred samples. Second, without normed measures in Add Health, we created health interference and parental closeness variables that may not reflect similar constructs in other studies. Finally, although several algorithms significantly predicted functioning in each functional domain, algorithms predicted no more than 5% variance in outcomes. Clearly many other variables, including SES, as well as neighborhood- and family-level variables are pertinent
to academic and health functioning (Hemingway, Nicholson, Stafford, Roberts, & Marmot, 1997; White, 1982). Finally, more rigorous ascertainment of ADHD and the use of normed or well-validated outcome measures is necessary to confirm these preliminary findings.

The current study calculated the base rates and total predictive value estimates of ADHD symptoms in the population-based Add Health sample. In addition, we examined the classification accuracy estimates of various ADHD symptom algorithms, as well as these algorithms’ associations with diverse measures of functioning. Whereas the two most sensitive algorithms (i.e., Sensitive and Classic) significantly predicted GPA, health interference, and parental closeness, the two most specific algorithms (i.e., Observed and Specific) did not significantly predict any functional domain. The Classic algorithm also yielded ADHD prevalence estimates that converged with recent meta-analytic findings, and was therefore the single most useful algorithm. Finally, the Clinical algorithm, constructed from symptom thresholds based on a case-control study, significantly predicted GPA and parental closeness.

The current study provided preliminary data on the prevalence and utility of ADHD symptoms and symptom algorithms in a population-based sample. We await additional future research that will develop and evaluate promising innovation in the assessment of ADHD across contexts.
Table 4-1. Demographic and clinical characteristics of the Add Health Sample at Wave 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>ADHD (n = 827)</th>
<th>Comparison (n = 10,420)</th>
<th>t/χ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>14.98 (1.70)</td>
<td>14.98 (1.70)</td>
<td>-.03</td>
</tr>
<tr>
<td>Number of Boys (%)</td>
<td>498 (60%)</td>
<td>4676 (45%)</td>
<td>70.85*</td>
</tr>
<tr>
<td>Number of non-White children (¹) (%)</td>
<td>284 (34%)</td>
<td>4726 (45%)</td>
<td>37.63*</td>
</tr>
<tr>
<td>Number of ADHD Inattention Symptoms</td>
<td>5.44 (2.36)</td>
<td>.78 (1.24)</td>
<td>56.24*</td>
</tr>
<tr>
<td>Number of ADHD Hyperactivity Symptoms</td>
<td>5.66 (1.81)</td>
<td>1.24 (1.39)</td>
<td>68.60*</td>
</tr>
<tr>
<td>Academic GPA (91% available data)</td>
<td>2.56 (.76)</td>
<td>2.84 (.74)</td>
<td>-10.28*</td>
</tr>
<tr>
<td>Health Interference (97% available data)</td>
<td>.85 (1.13)</td>
<td>.69 (.95)</td>
<td>4.06*</td>
</tr>
<tr>
<td>Parental Closeness (96% available data)</td>
<td>4.51 (.69)</td>
<td>4.63 (.54)</td>
<td>-5.06*</td>
</tr>
</tbody>
</table>

Note. ADHD = attention-deficit/hyperactivity disorder
* p < .05.
¹1,282 of 6,237 individuals who self-identified as White also self-identified as a member of another racial/ethnic group (21%).
Table 4-2. Base rates for ADHD symptoms\(^a\)

<table>
<thead>
<tr>
<th>Inattention Items(^b)</th>
<th>Base Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>You were easily distracted</td>
<td>.23</td>
</tr>
<tr>
<td>You failed to pay close attention to details or made careless mistakes in your work</td>
<td>.19</td>
</tr>
<tr>
<td>You were forgetful</td>
<td>.13</td>
</tr>
<tr>
<td>You had difficulty sustaining your attention in tasks or fun activities</td>
<td>.12</td>
</tr>
<tr>
<td>You lost things that were necessary for tasks and activities</td>
<td>.11</td>
</tr>
<tr>
<td>You had difficulty organizing tasks and activities</td>
<td>.09</td>
</tr>
<tr>
<td>You didn’t follow through on instructions and failed to finish work</td>
<td>.09</td>
</tr>
<tr>
<td>You didn’t listen when spoken to directly</td>
<td>.09</td>
</tr>
<tr>
<td>You avoided, disliked, or were reluctant to engage in work requiring sustained mental effort</td>
<td>.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hyperactivity/Impulsivity Items</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>You fidgeted with your hands or feet or squirmed in your seat</td>
<td>.32</td>
</tr>
<tr>
<td>You talked too much</td>
<td>.32</td>
</tr>
</tbody>
</table>
You felt “on the go” or “driven by a motor”  .23
You blurted out answers before the questions had been completed  .19
You felt restless  .14
You had difficulty doing fun things quietly  .13
You had difficulty awaiting your turn  .13
You left your seat in the classroom in other situations when being seated was expected  .08

Note. ADHD = attention-deficit/hyperactivity disorder

aSymptoms considered present when endorsed as “often” or “very often.”
bInattention items prefaced with the clause: “When you were between 5 and 12.”
Table 4-3. Total predictive power values (TPVs) for ADHD symptoms\(^a\)

<table>
<thead>
<tr>
<th>Inattention Items(^b)</th>
<th>TPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>You didn’t follow through on instructions and failed to finish work(^2)</td>
<td>.87</td>
</tr>
<tr>
<td>You were easily distracted(^3)</td>
<td>.87</td>
</tr>
<tr>
<td>You had difficulty sustaining your attention in tasks or fun activities(^3)</td>
<td>.86</td>
</tr>
<tr>
<td>You lost things that were necessary for tasks and activities(^3)</td>
<td>.86</td>
</tr>
<tr>
<td>You had difficulty organizing tasks and activities(^3)</td>
<td>.86</td>
</tr>
<tr>
<td>You avoided, disliked, or were reluctant to engage in work requiring sustained mental effort(^3)</td>
<td>.86</td>
</tr>
<tr>
<td>You were forgetful(^3)</td>
<td>.86</td>
</tr>
<tr>
<td>You failed to pay close attention to details or made careless mistakes in your work(^2)</td>
<td>.86</td>
</tr>
<tr>
<td>You didn’t listen when spoken to directly(^2)</td>
<td>.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hyperactivity/Impulsivity Items</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>You left your seat in the classroom in other situations when being seated was expected(^2)</td>
<td>.87</td>
</tr>
<tr>
<td>You had difficulty awaiting your turn(^3)</td>
<td>.86</td>
</tr>
</tbody>
</table>
You felt restless\(^3\) .86
You had difficulty doing fun things quietly\(^3\) .86
You blurted out answers before the questions had been completed\(^3\) .85
You felt “on the go” or “driven by a motor”\(^3\) .84
You fidgeted with your hands or feet or squirmed in your seat\(^3\) .83
You talked too much\(^3\) .81

Note. ADHD = attention-deficit/hyperactivity disorder
\(^1\)a rating of “sometimes,” “often,” or “very often” was most predictive of ADHD and considered a symptom endorsement.
\(^2\)a rating of “often” or “very often” was most predictive of ADHD and considered a symptom endorsement.
\(^3\)a rating of “very often” was most predictive of ADHD and considered a symptom endorsement.
\(^a\)Based on ADHD Classic method (≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “often” or “very often”).
\(^b\)Inattention items prefaced with the clause: “When you were between 5 and 12.”
Table 4-4. Number of individuals meeting ADHD symptom criteria according to each symptom algorithm\(^{ab}\)

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Algorithm</th>
<th>Observed</th>
<th>Sensitive</th>
<th>Specific</th>
<th>Clinical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
<td>Absent</td>
<td>Present</td>
<td>Absent</td>
<td>Present</td>
</tr>
<tr>
<td>ADHD (n = 827)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>190</td>
<td>637</td>
<td>827</td>
<td>0</td>
<td>159</td>
</tr>
<tr>
<td>Comparison (n = 10,420)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>10,420</td>
<td>4,829</td>
<td>5,591</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>190</td>
<td>11,057</td>
<td>5,656</td>
<td>5,591</td>
<td>159</td>
</tr>
</tbody>
</table>

**Note.** ADHD = attention-deficit/hyperactivity disorder; Observed = $\geq 6$ ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed at their optimal (i.e., highest total predictive value) level; Sensitive = $\geq 6$ ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “sometimes,” “often” or “very often”; Specific = $\geq 6$ ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “very often”; Clinical = $\geq 6$ ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed at the optimal (i.e., highest total predictive value) level ascertained in the case-control (i.e., clinical) study from Chapter 3.

\(^a\)Based on ADHD Classic method ($\geq 6$ ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “often” or “very often”).

\(^b\)Symptoms rated on the Retrospective ADHD questionnaire.
### Table 4-5. Classification accuracy of each ADHD symptom algorithm$^{ab}$

<table>
<thead>
<tr>
<th>Symptom Algorithm</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>.23</td>
<td>1.00</td>
<td>1.00</td>
<td>.94</td>
</tr>
<tr>
<td>Sensitive</td>
<td>1.00</td>
<td>.54</td>
<td>.15</td>
<td>1.00</td>
</tr>
<tr>
<td>Specific</td>
<td>.19</td>
<td>1.00</td>
<td>1.00</td>
<td>.94</td>
</tr>
<tr>
<td>Clinical</td>
<td>.74</td>
<td>.99</td>
<td>.91</td>
<td>.98</td>
</tr>
</tbody>
</table>

*Note.* ADHD = attention-deficit/hyperactivity disorder; PPV = positive predictive value; NPV = negative predictive value. Observed = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed at their optimal (i.e., highest total predictive value) level; Sensitive = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “sometimes,” “often” or “very often”; Specific = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “very often”; Clinical = ≥ 6 of 9 inattention and/or hyperactivity/impulsivity symptoms endorsed at the highest TPV level obtained from parent symptom ratings in the case-control (i.e., clinical) study from Chapter 3.  
$^a$Classification accuracy based on ADHD Classic method (≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “often” or “very often”). 
$^b$Symptoms rated on the Retrospective ADHD questionnaire.
Table 4-6. Associations of ADHD algorithms with functioning

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>R²</th>
<th>β</th>
<th>R²</th>
<th>β</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GPA</td>
<td></td>
<td>Health</td>
<td></td>
<td>Parents</td>
<td></td>
</tr>
<tr>
<td><strong>Symptom Algorithm</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classic</td>
<td>-7.19**</td>
<td>.03</td>
<td>3.43**</td>
<td>.02</td>
<td>-4.38**</td>
<td>.04</td>
</tr>
<tr>
<td>Observed</td>
<td>-2.41*</td>
<td>.02</td>
<td>2.48*</td>
<td>.02</td>
<td>-2.09*</td>
<td>.03</td>
</tr>
<tr>
<td>Sensitive</td>
<td>-12.18**</td>
<td>.05†</td>
<td>4.98**</td>
<td>.02†</td>
<td>-6.40**</td>
<td>.04†</td>
</tr>
<tr>
<td>Specific</td>
<td>-2.62*</td>
<td>.02</td>
<td>1.92</td>
<td>.02</td>
<td>-1.99*</td>
<td>.03</td>
</tr>
<tr>
<td>Clinical</td>
<td>-6.20**</td>
<td>.03</td>
<td>1.71</td>
<td>.02</td>
<td>-3.95**</td>
<td>.04</td>
</tr>
</tbody>
</table>

*Note. ADHD = attention-deficit/hyperactivity disorder; GPA = grade point average; Health = the frequency of a health or emotional problem having caused a missed day of school or social/recreational activity in the last month; Parents = average of how close one felt with their mothers and fathers, and whether their parents cared about them; Classic = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “often” or “very often”; Observed = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed at their optimal (i.e., highest total predictive value) level; Sensitive = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “sometimes,” “often” or “very often”; Specific = ≥ 6 ADHD inattention and/or hyperactivity/impulsivity symptoms endorsed as “very often”; Clinical = ≥ 6 of 9 inattention and/or hyperactivity/impulsivity symptoms endorsed at the highest TPV level obtained from parent symptom ratings in the case-control (i.e., clinical) study from Chapter 3.

*p ≤ .05.

**p ≤ .01 (α value based on Bonferroni correction).

†Highest $R^2$ value within a functional domain.
REFERENCES


doi:10.1097/00004583-199707000-00021

doi:10.1097/00004583-200010000-00002


doi:10.1097/00004583-198805000-00011


