10-2-2009

The Incremental Utility of Behavioral Rating Scales and a Structured Diagnostic Interview in the Assessment of ADHD

Aaron Vaughn
University of Vermont

THE INCREMENTAL UTILITY OF BEHAVIORAL RATING SCALES AND A STRUCTURED DIAGNOSTIC INTERVIEW IN THE ASSESSMENT OF ADHD

A Dissertation Presented

by

Aaron J. Vaughn

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
Specializing in Clinical Psychology

October, 2009

Accepted by the Faculty of the Graduate College, The University of Vermont, in partial fulfillment of the requirements for the degree of Doctor of Philosophy specializing in Clinical Psychology.

Thesis Examination Committee:

Betsy Hoza, Ph.D. **Advisor**

Rex Forehand, Ph.D.

Stephanie McConaughy, Ph.D.

Timothy Stickle, Ph.D.

David Rettew, M. D. **Chairperson**

Patricia A. Stokowski, Ph. D **Interim Dean, Graduate College**

Date: May 28, 2009

Abstract

Attention-Deficit/Hyperactivity Disorder (ADHD) is a disorder characterized by a persistent pattern of developmentally inappropriate levels of inattention, hyperactivity, and impulsivity (American Psychiatric Association, 2000). Currently, clinicians typically utilize a multi-method assessment battery focusing on identifying the core symptoms of ADHD. Further, current recommendations for a comprehensive assessment of ADHD require a lengthy and costly evaluation protocol despite a lack of evidence supporting the incremental utility of each method.

Assessment strategies exhibiting the strongest evidence of reliability and validity include symptom-based rating scales, empirically-derived rating scales, and structured diagnostic interviews (Pelham, Fabiano, & Massetti, 2005). However, limited empirical evidence supports their *incremental* validity in an assessment of ADHD. Prominent researchers have argued that labor-intensive measures such as a structured diagnostic interview provide only redundant information within an assessment that includes behavioral rating scales (Pelham et al., 2005), yet, their review provided limited empirical support for this conclusion. Nonetheless, other reviews have noted the lack of research examining whether each procedure and/or method adds unique information to a diagnosis of ADHD (Johnston & Murray, 2003).

In order to fill this gap in the literature, the current study examined the independent and incremental utility of multiple methods and informants in a comprehensive, "gold standard" assessment of ADHD. The sample included 185 children with ADHD (M_{age} = 9.22, SD =.95) and 82 children without ADHD (M_{age} = 9.24, SD = .88). Logistic regressions were used to examine the incremental contribution of each method in the prediction of consensus diagnoses derived by two Ph.D. level experts in the field of ADHD following a review of comprehensive assessment data. This study also examined the clinical utility and efficiency of diagnostic algorithms using the methods demonstrating the greatest statistical association with a diagnosis of ADHD.

Findings provided empirical support for arguments espousing the redundancy of information in a comprehensive assessment. Namely, information collected from a structured diagnostic interview was unable to significantly improve a prediction model including parent and teacher ratings (Block χ^2 = .91, ρ = .64). Importantly, parent and teacher ratings on a symptom-based scale alone were able to correctly classify 265 of 267 participants. Based on these results, a diagnostic algorithm that was derived utilizing only behavioral rating scales was able to classify correctly all 267 participants. Clinical implications are highlighted and future research directions are discussed.

Acknowledgments

Completion of an academic degree of this magnitude would have been impossible without the support, encouragement, and patience of countless mentors, advisors, friends, family members, and one beautiful, ever-enduring wife. To all those who guided and sustained me through these past six years, I cannot thank you enough. I would be remiss, however, not to express my appreciation towards those who without doubt were a necessary catalyst in the completion of this academic pursuit:

To my wife, Amanda Vaughn, you are the one most keenly aware of the sacrifices made in the pursuit of this degree. You have been steadfast in your support, patience, and willingness to face each challenge even when that meant leaving friends and family and moving so far away (and so ridiculously close to Canada). I believe with all of my heart that I would not be able to write these words today without you.

To my parents, Gary and Debby Vaughn, I can only imagine what it was like to have gone through this process with children. Every game of 1-2-3-4-5 in the front yard, trip to the park after school, and meal at the dinner table as a family is cherished even more after having a much greater understanding of the stress brought by a doctoral program. I pray that I will be able to provide the same amount of love and support to my child that you have given to me.

To my advisor, Betsy Hoza, in August of 2003, I started down this road having no idea where it would lead. In August of 2005, I took a turn onto a road leading 900 miles away from everything that I knew. Although I made these decisions with great fear and trepidation, they were unquestionably the best decisions I could have ever made. Few, if

any, would have so willingly given of their time, put forth as much effort, or re-read manuscripts as frequently in order to train one graduate so well (or so thoroughly). I hope that you are able to take pride and comfort in knowing that your wisdom, guidance, and meticulousness are both genuinely appreciated and deeply valued. I, with great sincerity and admiration, was and am incredibly lucky to have you as an advisor and mentor.

To the Hoza lab, Nina, Beth, Meg, Julia, and Kate, you put up with my inane questions and foolish mistakes as a naive first year through my never-ending "wisdom-givings" and persistent use of semi-useful (at best) checklists in my fifth year. Each of you showed me how to survive a 95-degree day locked in a room sorely lacking air conditioning while purposefully frustrating children and still be able come back joyfully the next day ready to do it all over again. I have few fonder memories than those of counselor training weeks where we each unabashedly revealed our true colors as stubborn, oppositional, 8-year-old children in a raucous, time-out riddled game of kickball. I am delighted to call you my friends, my colleagues, and, most importantly, my labmates.

To my committee, Drs. Rex Forehand, Stephanie McConaughy, Tim Stickle, and David Rettew, your insight, guidance, and support has been invaluable in completing this last, cumulative achievement. Further, each of you has had a profound impact on my research and clinical training. I have no doubt that each skill taught and refined through my training here at the University of Vermont will be of invaluable benefit throughout my career.

Table of Contents

Acknowledgements.....ii

List of Tables.....vi

List of Figures.....viii

Chapter 1: The Incremental Utility of Behavioral Rating Scales and a Structured Diagnostic Interview in the Assessment of Attention-Deficit/Hyperactivity Disorder

1.Introduction.....1

Current guidelines for the assessment of ADHD.....2

Current assessment strategies for ADHD.....4

Clinical utility of behavioral rating scales.....7

Clinical utility of structured diagnostic interviews.....10

Research examining the unique contribution of assessment instruments.....11

Study rationale and hypotheses.....13

2.Method.....17

Participants.....17

Consensus diagnosis.....17

Parent Measures.....19

Demographic Questionnaire.....19

Disruptive Behavior Disorders Rating Scale.....19

Child Behavior Checklist.....21

The Computerized Diagnostic Interview Schedule for Children-IV.....21

Teacher Measures.....23

Disruptive Behavior Disorders Rating Scale.....23

Teacher Report Form.....23

3.Data Analyses.....24

Preliminary analyses.....24

Primary analyses.....24

Statistical analyses.....24

Analyses examining the incremental utility of symptom-based and empirically-derived rating scales and a structured diagnostic interview 30

Analyses examining the incremental utility of parent and teacher- completed rating scales.....35

Analyses examining the incremental utility of parent or teacher completed rating scales and a structured diagnostic interview.....36

Examination of logistic regression models including significant covariates....41

Examination of the sensitivity, specificity, classification rates, and odds ratios within logistic regression models.....44

Examination of the validity and clinical utility of diagnostic algorithms.....46

4.Discussion.....51

Strengths and limitations.....57

Clinical and research implications.....58

References.....61

List of Tables

Table Page

Table 1: Consensus Diagnosticians' Subjective Weightings of Measures used in the Comprehensive Assessment of ADHD.....20

Table 2: Summary of Group Comparisons using Chi-Square Analyses for Categorical Variables.....25

Table 3: Summary of Univariate ANOVAs Comparing Children with and without ADHD on Measures of IQ, Internalizing and Externalizing Problems, and Age.....26

Table 4: Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Parent Ratings on the Disruptive Behavior Disorders Rating Scale and the Child Behavior Checklist.....32

Table 5: Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis Teacher Ratings on the Disruptive Behavior Disorders Rating Scale and the Teacher Report Form 33

Table 6: Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Parent Report on the Diagnostic Interview Schedule for Children-IV...34

Table 7: Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Parent and Teacher Ratings on the Disruptive Behavior Disorders Rating Scales, the Child Behavior Checklist, and the Teacher Report Form.....37

Table 8: Summary of Conditional Forward-Entry Logistic Regression Analysis Predicting Consensus Diagnosis using Parent and Teacher Ratings on the Disruptive Behavior Disorders Rating Scales.....38

Table 9: Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Parent Ratings on the Disruptive Behavior Disorders Rating Scale, the Child Behavior Checklist, and Parent Report on the Diagnostic Interview Schedule for Children-IV.....40

Table 10: Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Teacher Ratings on the Disruptive Behavior Disorders Rating Scale and Teacher Report Form and Parent Report on the Diagnostic Interview Schedule for Children-IV.....42

Table 11: Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Parent and Teacher Ratings on the Disruptive Behavior Disorders Rating Scales and Parent Report on the Diagnostic Interview Schedule for Children-IV.....43

Table 12: Summary of the Sensitivity, Specificity, Percent Agreement, and Chance-Corrected Agreement for Logistic Regression Models using Parent and Teacher Ratings on the Disruptive Behavior Disorders Rating Scales, Child Behavior Checklist, and Teacher Report Form, and Parent Report on the Diagnostic Interview Schedule for Children-IV.....45

Table 13: Summary of Sensitivity, Specificity, Percent Agreement, and Chance-Corrected Agreement for Diagnostic Algorithms utilizing Parent and Teacher Ratings on the Disruptive Behavior Disorders Rating Scales, the Child Behavior Checklist, and the Teacher Report Form.....48

List of Figures

Figure 1: Review of assessment methods including informants, administration procedures, and time requirements.....5

The Incremental Utility of Behavioral Rating Scales and a Structured Diagnostic Interview in the Assessment of Attention-Deficit/Hyperactivity Disorder

Attention-Deficit/Hyperactivity Disorder (ADHD) is a disorder characterized by a persistent pattern of developmentally inappropriate levels of inattention, hyperactivity, and impulsivity (American Psychiatric Association, 2000). As one might expect of children who are excessively hyperactive, distractible, and impulsive, children with ADHD exhibit difficulties in multiple areas of daily life functioning including (but not limited to) the academic, social, family, and behavioral domains (Pelham, Fabiano, & Massetti, 2005). Further, these impairments continue through adolescence and adulthood despite decreases in core symptoms (Barkley, Fischer, Smallish, & Fletcher, 2004). Past research examining the assessment of ADHD has focused on identifying these core symptoms in multiple settings for the purpose of establishing a diagnosis of ADHD (Pelham et al., 2005). Consequently, although diagnostic practices vary widely, clinicians typically utilize a multi-method assessment battery in the evaluation of ADHD that may include parent and child diagnostic interviews, behavioral rating scales completed by parents and/or teachers, direct observations, and/or clinic based assessments (e.g., continuous performance tasks) (Barkley, 1988a; DuPaul, 1991).

As a result of this approach, an abundance of research has considered the reliability and validity of numerous symptom-based rating scales, empirically-derived rating scales, and structured interviews in the assessment of ADHD; however, little research has examined the actual incremental validity and clinical utility of these methods as part of a comprehensive assessment of ADHD (DuPaul, 1991; Johnston & Murray,

2003; Pelham et al, 2005). In fact, a comprehensive assessment typically entails several hours of the clinician's, parent's, and child's time (Pelham et al, 2005). Given the cost and practicality of such an assessment, researchers have strongly advocated for studies identifying more efficient, cost effective assessments of ADHD (Johnston & Murray, 2003; Pelham et al., 2005). Thus, the purpose of the present study was to examine the incremental utility of symptom-based and empirically-derived rating scales and a structured diagnostic interview in the assessment of ADHD.

Current Guidelines for Assessment of ADHD

The clinical purpose of an ADHD assessment entails the need to establish an individual's need for treatment, derive appropriate treatment goals, and monitor the individual's progress and outcome (Pelham et al., 2005). Given the importance of accurately assessing and treating ADHD, the American Academy of Pediatrics' (AAP) current guidelines state that primary care physicians should (a) screen for ADHD when core symptoms are present, (b) employ DSM-IV-TR criteria, (c) gather information about DSM-IV-TR symptoms directly from parents and teachers, (d) assess for functional problems and coexisting conditions, and (e) not use other "diagnostic tests" (e.g., CPT) to confirm a diagnosis of ADHD (American Academy of Pediatrics; 2000, 2001).

Current DSM-IV-TR requirements include an assessment of functioning in multiple settings, documented age of onset and symptom criteria, and associated impairment to be eligible for a diagnosis (American Psychiatric Association, 2000). Currently, DSM-IV-TR criteria for a diagnosis include nine symptoms of inattention and/or nine symptoms of hyperactivity/impulsivity. An individual qualifies for a

diagnosis of ADHD Predominantly Inattentive Type if six or more symptoms of inattention only are reported, a diagnosis of ADHD Predominantly Hyperactive/Impulsive Type if six or more symptoms of hyperactivity/impulsivity only are reported, or a diagnosis of ADHD Combined Type if six or more symptoms are reported in each category. Research supports these recommendations as assessments utilizing DSM-IV criteria are more sensitive than other diagnostic approaches and facilitate appropriate treatment of individuals with ADHD (McGough & McCracken, 2000). Typically, a diagnosis is given by the clinician after reviewing the reports of individuals having extensive interactions with the child (parents and teachers). Consequently, when multiple raters in different settings agree on core symptoms, an ADHD diagnosis is straightforward (Wolraich, Lambert, Bickman, Simmons, Doffing, & Worley, 2004). However, assessment is complicated by low rates of cross-informant agreement and vague recommendations in the DSM-IV regarding incorporating discrepant information (Power, Costigan, Leff, Eiraldi, & Landau, 2001). To deal with these incongruities, current research strongly supports integrating information from a child's parent and teacher using behavioral symptom-based and empirically-derived rating scales and a structured interview for an appropriate assessment of ADHD (Pelham et al., 2005; Power, Doherty, Panichelli-Mindel, Karustis, Eiraldi, Anastopoulos et al., 1998; Tripp & Clarke, 2006). However, fulfilling these recommendations leads to an emphasis on a multi-informant, multi-method assessment that requires extensive cost, time, and resources with little evidence for the incremental validity of each step (Johnston & Murray, 2003). Accordingly, numerous researchers have argued for the need for more

efficient, more cost-effective assessment procedures (Johnston & Murray, 2003; Pelham et al., 2005).

Current Assessment Strategies for ADHD

As noted, the pervasive and chronic nature of problems associated with ADHD requires an evaluation that involves collecting data across multiple settings and caregivers. Current strategies utilized in the assessment of ADHD encompass five primary methods: structured diagnostic interviews, DSM-IV symptom-based rating scales, empirically-derived rating scales, direct observation, and continuous performance tasks (McGough & McCracken, 2000; Schaughency & Rothlind, 1991). Each method may require training in administrations, separate informants, and time and cost of scoring and interpretation (for a brief summary, see Figure 1). Given their use in clinical assessment (Pelham et al, 2005), we will review briefly the reliability, validity, and efficiency of each of these methods.

Structured diagnostic interviews have become increasingly common in the assessment of childhood behavioral disorders (Hodges, 1993). Though originally designed for epidemiological studies, their use in the clinical realm has increased due to their ability to assess a wide range of DSM-IV disorders in a consistent, standardized manner (McConaughy, 2000). However, extensive training is required prior to administration which itself typically requires a minimum of 45 to 60 minutes (though many administrations take considerably longer due to the severity/complexity of the psychopathology being reported) (McConaughy, 2000). Conversely, the ease and

Figure 1

Review of assessment methods including informants, administration procedures, and time requirements.

Method	**Informant**	**Administered By**	**Time Required (Efficiency)**
Structured Diagnostic Interviews (e.g., DISC)	Parent Child	Clinician Lay Interviewer	1 – 1.5 hours (may take longer given comorbidities)
Empirically-Derived Rating Scales (e.g., CBCL, BASC-2)	Parent Teacher Child self-report	Informant completed Clinician scored	10-15 minutes to complete; 10 minutes to score
DSM-IV Rating Scales (e.g., SNAP-IV, DBD)	Parent Teacher Child self-report	Informant completed Clinician scored	5-10 minutes to complete; 5 minutes to score
Direct Observation (e.g., Response Class Matrix)	Clinician Trained Observer	Observer completed Clinician scored	30 minutes to several hours
Continuous Performance Tests (e.g., GDS)	Child	Child Completed Clinician scored	30 minutes to administer; 10 minutes to score by computer

Note: DISC: Diagnostic Structured Interview for Children. CBCL: Child Behavior Checklist. BASC-2: Behavioral Assessment System for Children, 2nd Ed. SNAP-IV: Swanson, Nolan, and Pelham Rating Scale, 4th Ed. DBD: Disruptive Behavior Disorders Rating Scale. GDS: Gordon Diagnostic System.

efficiency of numerous symptom-based and empirically-derived rating scales has led to their frequent use in the assessment process (Mash & Terdal, 1997; Pelham et al, 2005). Rating scales allow clinicians to gather information efficiently (typically less than 20 minutes) from informants who have known the child for months or years (Mash & Terdal, 1997; Power, Andrews, Eiraldi, Doherty, Ikeda, DuPaul, & Landau, 1998). Direct observational procedures, although less efficient, provide opportunities for the examiner to record or code the child's behavior in a naturalistic (e.g., classroom) or laboratory setting. Numerous systematic observational procedures have been developed that show adequate reliability, interobserver agreement, and concurrent and discriminative validity with other assessment methodologies (Hintze & Matthew, 2004; Volpe, DiPerna, Hintze, & Shapiro, 2005). However, despite the quality of information gathered during direct observations, their lack of cost-effectiveness (lengthy observations over multiple time points), amount of training required to implement properly, and questionable ecological validity when laboratory settings are used limits the clinical utility/efficiency of observational methods in assessment. Continuous performance tasks (CPT) also have been used as an efficient assessment tool for ADHD (McCough & McCracken, 2000). Typically, CPTs require the child to emit a response (press button) when a specific stimulus is presented on a screen. Errors of omission (stimuli not responded to) and commission (incorrect stimuli responded to) are calculated to assess vigilance, sustained attention, and impulse control (Halperin, Sharma, Greenblatt, & Schwartz, 1991; Mash & Terdal, 1997). Despite reliably discriminating between children with ADHD and controls, studies have shown CPTs to have little ecological and

construct validity in the diagnosis of ADHD and to be poor at distinguishing between ADHD and other diagnostic classifications, producing unacceptable levels of false-positives (Barkley, 1991; Barkley & Grodzinsky, 1994; Matier-Sharma, Perachio, Newcorn, Sharma, & Halperin, 1995). Hence, their use in the diagnosis of ADHD is not recommended (American Academy of Pediatrics; 2000, 2001).

Due to the importance of valid, cost effective assessment for ADHD, an examination of the incremental contribution of each component of the assessment process is critical in providing the most efficient, treatment-informing assessment. Consequently, the current study examined assessment methods demonstrating the most promising evidence of efficiency and clinical utility: symptom-based rating scales, empirically-derived rating scales, and a structured diagnostic interview.

Clinical utility of behavioral rating scales. Behavioral rating scales of ADHD symptoms are an efficient and cost-effective tool for clinicians (Pelham et al., 2005; Owens & Hoza, 2003). Research has shown that parents and teachers are the optimal informants of disruptive behavior disorders and externalizing problems (Loeber et al., 1991). Further, studies have shown that combining parent and teacher ratings of ADHD symptoms increases the sensitivity (ability to identify a disorder when it is present) and specificity (ability to distinguish when a disorder is not present) of the measure beyond that provided by one informant (Power, et al., 1998a; Tripp, & Clarke, 2006). However, researchers have noted the surprising lack of research testing the unique or additional information provided by different informants as part of the assessment process as clinicians tend to have "blind faith in the 'more is better' approach" (p. 500; Johnston &

Murray, 2003). Further, other studies have reported low rates of specificity, high rates of cross-informant discrepancies (e.g., raters from different settings), and low rates of interobserver agreement (e.g., raters from the same setting) in evaluations utilizing behavioral rating scales (Collett, Ohan, & Myers, 2003; Reid & Maag, 1994; Wender, 2004). Consequently, researchers have suggested a multiple-gating approach utilizing the efficiency and positive predictive power of rating scales as a screener to identify children in need of further assessment, though this has not been implemented in common practice (Simonsen & Bullis, 2007; Tripp & Clarke, 2006).

In particular, DSM-IV symptom-based rating scales provide information about behavior in multiple settings (e.g., home and school), discriminate between clinical and nonclinical groups and between subtypes of ADHD, and are sensitive to both behavioral and pharmacological treatment effects (MTA Cooperative Group, 1999; Power et al., 1998a; Power et al., 1998b; Tripp & Clarke, 2006). These measures utilize each of the 18 symptoms defining ADHD on a Likert scale and typically have norms and/or suggested cutpoints for a diagnosis of ADHD. Several similar well-validated scales have been utilized in the assessment of ADHD, including: the Swanson, Nolan, and Pelham-IV Rating Scale (Atkins et al., 1985; Swanson, 1992); the Disruptive Behavior Disorders Rating Scale (Massetti et al., 2003; Pelham, Gnagy, Greenslade, & Milich, 1992); the ADHD Rating Scale-IV (DuPaul, 1991; DuPaul, Power, Anastopoulos, & Reid,1998); and the Vanderbilt Rating Scale (Wolraich, Feurer, Hannah, Baumgaerel, & Pinnock, 1998). However, researchers have argued that the common use of fixed cutoff scores to determine the presence of ADHD symptoms for the purpose of diagnosing ADHD may

not be an optimal strategy (Eiraldi, Power, Karustis, & Goldstein, 2000; Power et al., 2001). Specifically, the majority of these measures implicitly assume that symptoms noted as frequently occurring also include accompanying functional impairment (Reid & Magg, 1994). Further, symptoms of ADHD (the primary targets of DSM-based rating scales) are not robust predictors of long-term outcomes and often are not the reason for referral for treatment (Angold, Costello, Farmer, Burns, & Erkanli, 1999; Manuzza & Klein, 1999; Pelham et al., 2005).

Empirically-derived rating scales also have been shown to accurately identify children with ADHD, to discriminate children with inattention only from children with inattention and hyperactivity, and to be sensitive to treatment effects (for a review see Pelham et al., 2005). Further, these measures typically provide norm-referenced information across age and gender (Achenbach & Rescorla, 2001; Hartman, Stage, & Webster-Stratton, 2003). Of particular interest to this study, the Attention Problems Syndrome Scales on the Child Behavior Checklist (CBCL) and Teacher Report Form (TRF) have been used as a proxy for diagnosis in several studies, are highly related to DSM-IV ADHD diagnoses, are able to discriminate between subtypes of ADHD, are sensitive to both behavioral and treatment effects, and have been used to assess outcomes in numerous treatment studies (Barkley, Shelton, Crosswait, Moorehouse, Fletcher, Barrett, et al., 2000; Chen, Faraone, Biederman, & Tsuang, 1994; Hartman, et al., 2003). Although these findings provide support for the inclusion of empirically-derived rating scales in the assessment process, little research has examined their incremental contribution in the assessment of ADHD.

Clinical utility of structured diagnostic interviews. In order to increase the reliability and validity of the diagnostic process, assessments often utilize structured diagnostic interviews (McConaughy, 2000). Structured diagnostic interviews utilize standardized questions to reduce the variability in responses while maintaining the ability to assess for multiple disorders within the same evaluation (Pelham et al., 2005; Schaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000). This structure was designed to reduce clinicians' tendencies to collect information selectively (leading to an incorrect diagnosis), make diagnoses most familiar to them, or determine a diagnosis before all relevant information is collected (McClellan & Werry, 2000). Researchers have shown that structured diagnostic interviews exhibit satisfactory rates of sensitivity and specificity, convergent and discriminant validity, and concurrent validity in relation to DSM-IV diagnostic categories (Pelham et al., 2005). Further, clinicians have utilized structured interviews as an indicator of initial diagnosis, treatment response, and as a comparison for prior diagnoses utilizing other strategies (Piacentini, Roper, Jensen, et al., 1999). However, researchers have noted significant decreases in symptom endorsement over the course of administration (attenuation effects), influence of interviewer and subject characteristics on diagnosis, and low to moderate levels of test-retest reliability (Jensen & Edelbrock, 1999; Piacentini et al., 1999; Roberts, Solovitz, Chen, & Casat, 1996). Further, structured diagnostic interviews require considerable time to administer, score, and interpret. For example, the Diagnostic Interview Schedule for Children-IV (DISC-IV) is composed of 358 "stem questions" and almost 3,000 additional questions that may be administered (Shaffer et al., 2000). Shaffer and colleagues (2000) reported

an average of 70 minutes per administration per informant. Consequently, researchers have questioned the efficiency of structured diagnostic interviews given the ease and speed with which behavioral rating scales may be administered and scored (Pelham et al., 2005). Indeed, Pelham and colleagues (2005) argue that structured diagnostic interviews provide little information beyond that gathered through more efficient methods (e.g., behavioral rating scales) in the assessment of ADHD.

Research Examining the Unique Contribution of Assessment Instruments

A recent review by Pelham and colleagues (2005) examined the separate components commonly used in the assessment of ADHD. Their review provided strong support for the reliability and validity of each of the above assessment techniques independently. Specifically, they noted that DSM-IV symptom-based and empirically-derived behavioral rating scales, structured diagnostic interviews, observational measures, and other less frequently used methods (e.g., impairment measures) all exhibit appropriate reliability and validity (e.g., concurrent, convergent, and discriminant) in the assessment of ADHD. However, they noted the lack of research examining the clinical utility of each of these methods from an incremental standpoint. In fact, they argued for the use of "the minimum strategies and tools necessary for an efficient and effective assessment for ADHD" (p. 465; Pelham et al., 2005). In support of their conclusion, they reviewed recent research supporting the efficiency and proposed utility of single raters, shorter empirically-derived scales, and/or subsets or even individual items in identifying children with ADHD. However, other studies have reported differing positive and negative predictive power of individual symptoms of ADHD when rated by either a

child's parent, teacher, or both (Power, et al, 1998a; Power, et al., 1998b; Wolraich, et al., 2004). One study examining assessments utilizing parent- and teacher-rated symptoms of ADHD concluded that the optimal approach varied as a function of informant, scale, and purpose of assessment (Power et al., 2001). However, their study utilized a specific set of screening rules and diagnostic procedures that included a structured diagnostic interview, but did not examine the incremental validity of parent and teacher report in conjunction with the structured diagnostic interview. Hence, despite the wealth of evidence supporting the differing predictive power of ADHD symptoms by method and by informant, few studies have examined the incremental contribution of different types of scales, structured diagnostic interviews, and/or informants thereby failing to answer fully this particular question of clinical utility and efficiency (Pelham et al., 2005).

Further, despite the reliability and validity exhibited by structured diagnostic interviews, Pelham and colleagues (2005) argued that they demonstrate little incremental utility within a comprehensive assessment that also includes parent and teacher behavioral rating scales. Their review notes the high correlation between structured diagnostic interviews and rating scales and the near perfect rates of agreement when classifying children using rating scales *or* a structured diagnostic interview (DuPaul, Power, McGoey, Ikeda, & Anastopoulos, 1998; Ostrander, Weinfurt, Yarnold, & August, 1998; Power et al., 2001). Following this line of research, they argue that "Diagnosing ADHD is most efficiently accomplished with parent and teacher rating scales" (p. 469; Pelham et al., 2005). However, scant research has directly compared the incremental utility of rating scales *and* a structured diagnostic interview in the assessment of ADHD.

In fact, past research has either 1) compared the utility of different raters in predicting a diagnosis of ADHD or 2) examined the correlations among behavioral rating scales and structured diagnostic interviews. Further, despite the lack of empirical support for the incremental utility of multiple assessment methods, current recommendations for a comprehensive assessment of ADHD require a lengthy and costly evaluation protocol (Johnston & Murray, 2003; Tripp & Clark, 2006). We argue that research must not only examine the degree of correlation among multiple methods and raters but also examine the actual, independent contribution each adds in the assessment of ADHD. Only studies utilizing this type of design are able to inform the relation between incremental utility and cost-effectiveness (Johnston & Murray, 2003). Consequently, the present study examines the independent *and* incremental contributions provided by various informants, rating measures, and a structured diagnostic interview in the comprehensive assessment of ADHD.

Study Rationale and Hypotheses

Though studies have examined the predictive, and to a lesser extent, the clinical utility of parent and teacher rating scales *or* structured diagnostic interviews *separately*, there is a paucity of research examining the incremental utility of these methods together. Further, the overwhelming consensus in the field is to use a comprehensive assessment entailing multiple informants utilizing multiple methods despite recent arguments suggesting that this "more is better" approach may be unwarranted (Pelham et al., 2005). Consequently, the purposes of this study were threefold: 1) to examine the *independent* predictive utility of symptom-based rating scales, empirically-derived rating scales, and a

structured diagnostic interview within a multi-method, multi-informant assessment of ADHD; 2) to examine the *incremental* predictive utility of each method and informant in the prediction of diagnostic status (e.g., ADHD or Not ADHD); and 3) to examine the *incremental* utility and *efficiency* of clinically-relevant algorithms using the methods demonstrating the greatest association with a diagnosis of ADHD.

Given the extensive recommendations for the inclusion of each of the reviewed methods in the assessment of ADHD (McGough & McCracken, 2000; Wolraich, et al., 2004), this study was designed to evaluate how diagnostic prediction is *improved* through the inclusion of each method and/or informant. Consistent with our first goal, we first examined all methods simultaneously in order to clarify the *unique* information contributed by each method in the prediction of a diagnosis of ADHD. Subsequently, we then examined the incremental contributions of each measure within each method in the prediction of a diagnosis of ADHD. Consistent with previous research, we hypothesized that each method individually would be significantly associated with a diagnosis of ADHD (Chen et al., 1994; Power et al., 1998a; Power et al., 1998b; Schaffer et al., 2000). In addition, given the strength of the reviewed evidence for the reliability and validity of symptom-based and empirically-derived rating scales (Chen et al., 1994; Pelham, et al., 2005; Powers, et al., 2001;), we hypothesized that models employing both symptom-based and empirically-derived rating scales would account for significantly greater variance than models utilizing symptom-based ratings alone. This procedure of entering methods into statistical models in order of their relative efficiency is consistent with our goal of examining the incremental utility *and* efficiency of each method.

Second, we examined the incremental utility of each method *and* rater in the prediction of diagnostic status. Given the evidence supporting the increased predictive validity of models including multiple informants (e.g., parent and teacher) (Power, et al., 2001; Power et al., 1998a), we hypothesized that teacher ratings of ADHD symptoms, when examined incrementally with parent ratings, would result in significantly improved predictive models. Further, despite near universal endorsement for their inclusion in the assessment of ADHD, little, if any, empirical evidence exists to support the inclusion of structured diagnostic interviews within an assessment of ADHD that also includes behavioral rating scales. As such, to our knowledge, this is the *first* study to examine the incremental utility of a structured diagnostic interview *beyond that* of the more efficient rating scales. We posited that a structured diagnostic interview would not account for significant variance beyond that accounted for by the more efficient parent *and* teacher rating scales. Such a result would provide empirical support for the conclusion of Pelham and colleagues (2005) that these methods provide redundant information. Further, given the redundancy in informants utilizing parent-completed rating scales and a structured diagnostic interview (also completed by the parent), we hypothesized that a structured diagnostic interview would not significantly improve a model including parent-completed rating scales. Alternatively, we hypothesized that a structured diagnostic interview would significantly improve a model including teacher-completed rating scales given the non-redundancy in informants.

Finally, we examined the clinical utility of diagnostic algorithms that utilized the methods examined above. As argued by Johnston and Murray (2003), research regarding

incremental validity "...must be conducted with procedures, measures, and samples that reflect the realities of clinical practice" (p. 504). As such, we wished to explore the utility and efficiency of various clinically-relevant diagnostic algorithms (of increasing complexity) in classifying children when compared to a diagnosis of ADHD derived using a comprehensive "gold standard" assessment. Given the lack of expediency in employing statistical models in a clinic setting, these algorithms may provide practical, real-world procedures to utilize *most efficiently* the various methods informing the assessment of ADHD.

Given current controversy in the literature regarding whether ADHD, Predominantly Inattentive Type (ADHD-I) should be considered a subtype of ADHD given the unique etiology, core deficits, associated features, and comorbid functioning of children diagnosed with ADHD-I (Milich, Balentine, & Lynam, 2001), only children with either ADHD, Combined Type (ADHD-C) or ADHD, Predominantly Hyperactive/Impulsive Type (ADHD-HI) were examined in this study. Given this selection bias, we expected that ratings of hyperactivity/impulsivity would account for greater variance than ratings of inattention in predicted models.

Method

Participants

Participants were drawn from an existing database from a study funded by the National Institute of Mental Health addressing unrelated research questions. Data collection occurred at three different sites including a large, public Midwestern university and two large, public Northeastern universities. The original purpose of the study was to examine children's self-perceptions in relation to their behavior. Participants were children aged 7 to 11 years of age. Recruitment was designed to gain a representative sample through the use of school settings, primary medical care settings, mental health practitioners, and self-referrals solicited through advertisements and word of mouth. Participants who met criteria for ADHD-C or ADHD-HI following the assessment procedures discussed below were placed in the ADHD group ($n = 185$). Children not meeting criteria for ADHD were placed in the control group ($n = 82$). Children with a Brief Intellectual Ability standard score below 80 on the Woodcock-Johnson Test of Cognitive Abilities (WJ-III) (Woodcock, McGrew, & Mather, 2001), a previous diagnosis of any pervasive developmental disorder, or who were currently taking medications that affected behavior and that could not be withdrawn for testing were excluded from the study. As noted above, children meeting criteria for ADHD-I also were ineligible for the study.

Consensus diagnosis. All participants received a comprehensive assessment of ADHD that followed current "gold standard" guidelines as established by the American Academy of Pediatrics (2000) and practice parameters as outlined by the American

Academy of Child and Adolescent Psychiatry (Pliszka, Bernet,, Bukstein, & Walter, 2007). Specifically, information was gathered through the use of parent and teacher symptom-based and empirically-derived rating scales, a comprehensive structured diagnostic interview, and a semi-structured clinical interview (regarding developmental, social, academic, and family functioning). Current functioning and impairment were assessed across settings including onset, frequency, and intensity of any behavioral, emotional, or academic concerns. Each child received an assessment battery including: 1) a cognitive and achievement battery; 2) self-report measures of self-perception, anxiety symptoms, and depressive symptoms; and 3) and a clinical interview. Clinicians administering the assessment were licensed psychologists or trained graduate-level research assistants who were supervised by licensed clinical psychologists.

Diagnostic decisions were made by licensed, Ph.D. level clinical psychologists specializing in ADHD and childhood Disruptive Behavior Disorders (hereafter referred to as the consensus diagnosticians). Each participant's comprehensive assessment data was reviewed by two independent consensus diagnosticians who derived a diagnosis for each child. The assessment data reviewed included the semi-structured clinical interview; parent- and teacher- completed Disruptive Behavior Disorders Rating Scales (DBD) (Pelham et al., 1992); the Computerized Diagnostic Interview Schedule for Children-IV parent version (DISC-IV) (Shaffer et al., 2000); the Child Behavior Checklist (CBCL) and Teacher Report Form (TRF) (Achenbach & Rescorla, 2001); and the Woodcock-Johnson III Tests of Cognitive Abilities (WJ-TC) and Achievement (WJ-TA) (Woodcock, et al., 2001). The Test Observation Form (TOF) (McConaughy &

Achenbach, 2004) also was available to the clinicians for a subset of children (n = 86) at one site. When diagnostic decisions were not in agreement, the consensus diagnosticians discussed the participant's assessment data until a consensus diagnosis was agreed upon (a total of 5 diagnosticians participated in the study). Agreement could not be reached with only one participant who was summarily excluded from further analyses. Further, to ascertain typical use of assessment data, each consensus diagnostician independently provided their subjective weightings of each primary assessment measure utilized in this study to examine for any possible procedural deviations or overreliance on a particular measure or method. Not surprisingly, weightings were consistent across consensus diagnosticians with no measures weighted less than 5% or greater than 25% with average weightings ranging from 7.6% to 16.4% (Table 1).

Parent Measures

Demographic Questionnaire (DQ). Demographic information including parental income, educational level, occupation, and marital status was provided by each participant's caretaker.

Disruptive Behavior Disorders (DBD) Rating Scale (Massetti, et al., 2003: Pelham, et al., 1992). The DSM-IV version of the parent DBD rating scale is a measure of the DSM-IV-TR (American Psychiatric Association, 2000) symptoms of ADHD, Oppositional Defiant Disorder (ODD), and Conduct Disorder (CD). The measure also includes some DSM-III-R (APA, 1987) symptoms of ADHD. Each of the 45 items is rated by the parent on a 4-point scale ranging from 0 (not at all present) to 3 (very much present). Items rated as a 2 (pretty much present) or 3 (very much present) are

Table 1

Consensus Diagnosticians' Subjective Weightings of Measures used in the Comprehensive Assessment of ADHD

Measure - Scale	**Diagnostician 1**	**Diagnostician 2**	**Diagnostician 3**	**Diagnostician 4**	**Diagnostician 5**	**Average Rating**
PDBD-IA	10	15	15	12	10	12.4
PDBD-HI	10	15	15	12	10	12.4
TDBD-IA	25	15	15	12	15	16.4
TDBD-HI	25	15	15	12	15	16.4
CBCL-A	5	8	5	10	10	7.6
TRF-A	5	8	5	10	10	7.6
DISC-IA	10	12	15	16	15	14.6
DISC-HI	10	12	15	16	15	14.6
Total	100	100	100	100	100	

Note: PDBD-IA: Parent Disruptive Behavior Disorder Rating Scale – Inattention Subscale. PDBD-HI: Parent Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. TDBD-IA: Teacher Disruptive Behavior Disorders Rating Scale – Inattention Subscale. TDBD-HI: Teacher Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. CBCL-A: Child Behavior Checklist – Attention Problems Syndrome Scale. TRF-A: Teacher Report Form – Attention Problems Syndrome Scale. DISC-IA: Diagnostic Interview Schedule for Children-IV – Inattention Module. DISC-HI: Diagnostic Interview Schedule for Children-IV – Hyperactivity/Impulsivity Module.

considered to be endorsed as symptoms. Cronbach's alphas of .72 to .95 for all subscales have been reported for a predominantly school sample of children with and without ADHD symptoms (Owens & Hoza, 2003). For the present study, only the nine inattention and nine hyperactivity/impulsivity symptoms from the DSM-IV version of the DBD were used (comprising the inattention and hyperactivity/impulsivity scales, respectively). The coefficient alphas for the present sample were .93 for the inattention scale and .96 hyperactivity/impulsivity scale.

Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001). The CBCL is a 118-item parent-completed behavior problem checklist utilizing a Likert scale designed to assess multiple domains of children's externalizing and internalizing functioning in the past six months. The CBCL has been used widely for obtaining ratings of problem behavior in children and has demonstrated strong evidence of reliability and validity (Achenbach & Rescorla, 2001). Syndrome scales on the CBCL were derived through factor analyses. For this study, only the Attention Problems Syndrome Scale was used. An alpha coefficient of .86 and a test-retest reliability of .92 has been reported for the Attention Problems Syndrome Scale (Achenbach & Rescorla, 2001). The coefficient alpha for the present sample was .89.

The Computerized Diagnostic Interview Schedule for Children-IV (DISC-IV; Costello, Edelbrock, & Costello, 1985; Shaffer, et al., 2000). The DISC-IV is a structured diagnostic interview designed for use by lay interviewers in epidemiological studies to elicit DSM-IV-TR and ICD-10 diagnoses for children and adolescents covering 36 mental health disorders (Shaffer et al, 2000). There are 358 "stem" questions that are

asked of every respondent that are overly sensitive in order to lead to more "contingent" questions that are able to differentiate true positives from false positives. This study utilized the ADHD module in the computerized version of the DISC-IV. Shaffer et al.'s (2000) review of the DISC-IV found acceptable test-retest reliability ($\kappa = .79$) over a one-year period for the ADHD module in a clinical sample and a kappa of .60 and intraclass correlation of .84 for the test-retest reliability and symptom counts of the ADHD module for a previous version of the DISC-IV (DISC-2.3) in a community sample. Research has supported the concurrent criterion validity of the ADHD module in relation to other diagnostic interviews, symptom checklists, and external validators (e.g., school dysfunction, functional impairment) (Cohen, O'Conner, Lews, Velez, & Malachowski, 1987; Jensen, Koretz, Locke, et al., 1993; Jensen, Wantanabe, Richters, et al., 1996; Piacentini, Shagger, Fisher, Schwab-Stone, Davies, & Gioia, 1993;). For a symptom to be endorsed as present, criterion questions examine the relative frequency and duration of the symptom in multiple settings. Specifically, a symptom must be reported as present at home and/or at school for a period of at least six months. Questions are also addressed regarding initial onset of ADHD symptoms (e.g., prior to age seven) and whether symptoms have remitted for substantial period of time (i.e., longer than two months). Follow-up contingent questions also review associated impairment; however, as these items are not utilized by the standard DISC-IV diagnostic algorithm, they were not included in this study. The number of symptoms identified as present on the DISC-IV and the diagnostic classifications (e.g., ADHD or Not ADHD) were each examined in the

study. The coefficient alphas for the present sample were .93 for the inattention scale and .90 hyperactivity/impulsivity scale.

Teacher Measures

DBD Rating Scale (Massetti, et al., 2003; Pelham, et al., 1992). The teacher DBD Rating Scale was identical to the parent DBD Rating Scale. The Inattention and Hyperactivity/Impulsivity subscales also were used from the teacher version. The coefficient alphas for the present sample were .95 and .95 for the inattention and hyperactivity scales, respectively.

Teacher Report Form (TRF; Achenbach & Rescorla, 2001). The TRF is a 118-item teacher-completed behavior problem checklist utilizing a Likert scale to assess multiple domains of children's internalizing and externalizing functioning in the past six months. The TRF has strong evidence of reliability and validity and has been widely used for obtaining ratings of problem behavior in children (Achenbach & Rescorla, 2001). For this study, only the empirically-derived narrow-band Attention Problems Syndrome Scale was used. A coefficient alpha of .95 and test-retest reliability of .95 has been reported for the Attention Problems Syndrome Scale (Achenbach & Rescorla, 2001). The coefficient alpha for the present sample was .87.

Data Analyses

Preliminary Analyses

Preliminary analyses were conducted comparing groups (e.g., ADHD or Not ADHD) with regard to sex, a diagnosis of ODD/CD, grade retention, race/ethnicity, cognitive ability, Internalizing and Externalizing Problems scores from the CBCL and TRF, and age. Categorical variables were compared using chi-square analyses and univariate ANOVAs were conducted for continuous variables. Chi-square analyses indicated that the children in the ADHD group were more likely to be male, had significantly higher rates of ODD or CD diagnoses, and were significantly more likely to have been retained than children in the control group; no group differences were found related to race/ethnicity (Table 2). Univariate ANOVA analyses indicated that the ADHD group had significantly lower standard scores on the Brief Intellectual Ability scale of the WJ-III and significantly higher rates of Internalizing and Externalizing Problems on both the CBCL and TRF (as indexed by T-scores) compared to the control group (Table 3).

Primary Analyses

Statistical Analyses. The primary data analytic strategy employed was logistic regression. The binary dependent variable was consensus diagnosis (i.e., ADHD or Not ADHD). The primary independent variables were the inattention and hyperactivity/impulsivity subscales of the parent and teacher DBD (PDBD-IA, PDBD-HI, TDBD-IA, and TDBD-HI), the Attention Problems Syndrome Scales from the CBCL

Table 2

Summary of Group Comparisons using Chi-Square Analyses for Categorical Variables

	Control Group	**ADHD Group**	χ^2	**p**
Sex (% Male)	69.5%	81.1%	4.37	.037
ODD/CD Diagnosis (% with Diagnosis)	1.2%	66.5%	97.43	<.001
Grade Retained (% Retained)	2.4%	14.6%	8.88	.003
Race (% White)	76.8%	82.1%	5.52	.238

Table 3

Summary of Univariate ANOVAs Comparing Children with and without ADHD on Measures of IQ, Internalizing and Externalizing Problems, and Age

	Control Group		ADHD Group			
	Mean	**Standard Deviation**	**Mean**	**Standard Deviation**	**F**	**ρ**
BIA	107.89	14.88	97.02	14.71	30.83	<.001
CBCL-Int	49.79	10.31	61.31	10.13	72.66	<.001
CBCL-Ext	45.88	9.10	64.27	9.06	233.55	<.001
TRF-Int	47.27	7.65	59.51	9.51	105.57	<.001
TRF-Ext	47.94	7.06	62.96	8.56	194.10	<.001
Age	9.25	.88	9.22	.95	.09	.766

Note: BIA: Brief Intellectual Ability Standard Score as indexed on the Woodcock-Johnson III Tests of Cognitive Abilities. CBCL-Int: Child Behavior Checklist – Internalizing Problems Broadband Scale. CBCL-Ext: Child Behavior Checklist – Externalizing Problems Broadband Scale. TRF-Int: Teacher Report Form – Internalizing Problems Broadband Scale. TRF-Ext: Teacher Report Form – Externalizing Problems Broadband Scale.

and TRF (CBCL-A and TRF-A), and the number of parent-reported symptoms of inattention and hyperactivity/impulsivity on the DISC-IV (DISC-IA and DISC-HI). Of note, the statistical utility of mean DBD subscale scores relative to the number of symptoms endorsed as present (e.g., "Pretty Much" or "Very Much") was compared on each subscale; results indicated that models utilizing mean scores yielded greater overall fit, greater association, and greater classification. Therefore, only models utilizing the mean ratings of inattention and hyperactivity/impulsivity on the DBD scales are reported here. Likewise, comparisons of the number of symptoms endorsed on the DISC-IA and DISC-HI scales versus categorical diagnostic classifications indicated that models utilizing the number of parent-reported symptoms resulted in greater fit, greater association, and greater classification. Hence, only models utilizing the number of parent-reported symptoms on the DISC-IA and DISC-HI scales are reported here. Finally, T-scores on the CBCL-A and TRF-A were used in the logistic regression models.

Initially, to ascertain the *unique* variance accounted for by each method, all variables were examined *simultaneously* within a logistic regression model. However, this model resulted in complete separation of the data (Albert & Anderson, 1984); in other words, a linear function of our predictors generated perfect predictions of consensus diagnosis. This resulted in a failure of the likelihood maximization algorithm to converge within the regression model. Consequently, for this regression equation, maximum likelihood estimates simply did not exist (Allison, 2008; Albert & Anderson, 1984). As such, we were unable to examine the full model. Further, as seen in Table 7 (pg. 37), a logistic regression model including parent and teacher ratings resulted in extremely large

odds ratios and insignificant Wald statistics each of which is strongly indicative of quasi-complete separation (Heinz & Schemper, 2002). Indeed, the model closely approximated perfect prediction as 99.3% of all cases were classified correctly using the predicted model. As such, the instability of the maximum likelihood estimates, resulting logit coefficients, and inflated odds ratios were rendered uninterpretable (despite the significance of the overall model). Consequently, instead of the proposed "deconstructive" approach to inform examination of individual methods, our analyses followed a model building approach using *hierarchical* logistic regression models. Specifically, we examined the observed increase in model prediction (Block χ^2) and classification (percent of cases classified correctly) following the addition of each method within the logistic regression models.

Within each logistic regression model, results were examined using three different chi-square tests and a measure of association. Specifically, the Likelihood Ratio Chi-Square test was utilized to test the significance of the entire model (Model χ^2) and the significance of each block/step (Block χ^2) within the model (indicating that the block effect significantly improved the overall model). The Hosmer and Lemeshow Chi-Square test was utilized to examine the fit of the predicted model (nonsignificant ρ values indicate failure to reject the null hypothesis that there is no difference between the observed and predicted models) (Hosmer & Lemeshow, 2000); whereas, Nagelkerke's R^2 was used to examine the relative strength of association between the independent and dependent variables. We also examined the percentage of participants correctly classified using the predicted model. Although reported within the tables, the Wald statistic was

not examined directly. This was due to the presence of large logit coefficients and inflated standard errors within the regression equation which increases the probability for Type II errors when utilizing the Wald statistic (i.e., the squared ratio of the unstandardized logit coefficient to its standard error). This statistical flaw results in insignificant Wald Chi-Square values despite large effects (Agresti, 1996; Menard, 2002).

Consistent with the goals of the study, initial analyses utilized three logistic regressions to examine: 1) the independent and incremental utility of symptom-based and empirically-derived, *parent*-completed rating scales; 2) the independent and incremental utility of symptom-based and empirically-derived, *teacher*-completed rating scales; and 3) the independent and incremental utility of the inattention and hyperactivity/impulsivity scales from the structured diagnostic interview.

Following this series of logistic regression models, a second series of logistic regression models was conducted to examine the incremental utility of multiple methods *and* multiple informants. As outlined in the goals and hypotheses, these analyses examined the incremental utility of multiple informants (parent and teacher) and the addition of a structured diagnostic interview to the models outlined in the first and second logistic regression models outlined above. Specifically, we examined four logistic regressions: 1) the independent and incremental utility of teacher ratings in a model including parent ratings; 2) the independent and incremental utility of a structured diagnostic interview in a model including parent ratings; 3) the independent and incremental utility of a structured diagnostic interview in a model including teacher

ratings; and 4) the independent and incremental utility of a structured diagnostic interview in a model including *both* parent *and* teacher symptom-based ratings. As reviewed earlier, a model including all methods was not examined as maximum likelihood estimates could not be computed.

Finally, using the above data to maximize clinical utility and efficiency, we derived clinically-relevant diagnostic algorithms that utilized the measures, informants, and methods demonstrating the greatest incremental utility in the prediction of consensus diagnosis. The sensitivity, specificity, chance-corrected rate of agreement, and simple percent agreement of each algorithm was then examined in relation to the "gold-standard" consensus diagnosis.

Analyses Examining the Incremental Utility of Symptom-Based and Empirically-Derived Rating Scales and a Structured Diagnostic Interview

The initial hierarchical logistic regression examined the contribution of the parent-completed rating scales in the prediction of a consensus diagnosis of ADHD. Given the efficiency of symptom-based rating scales, these were entered into the model first followed by ratings on the empirically-derived scale to examine their relative incremental utility in the prediction of consensus diagnosis. In particular, we hypothesized that symptom-based (Inattention and Hyperactivity/ Impulsivity subscales on the DBD) and empirically-derived (Attention Problems syndrome scale on the CBCL) rating scales each would be associated with consensus diagnosis when examined simultaneously.

As seen at step 1 in Table 4, parent ratings of inattention *and* hyperactivity/ impulsivity on the DBD each contributed significant unique variance when entered simultaneously in the prediction of consensus diagnosis. Analyses also indicated that the incremental contribution of the CBCL-A at step 2 significantly improved a model including the PDBD-IA and PDBD-HI scales (Block $\chi^2 = 6.37$, $\rho < .05$). However, this model also indicated that only the PDBD-HI and CBCL-A were significantly associated with consensus diagnosis when all three predictors were included in the model (Table 4).

The second logistic regression model examined the incremental utility of teacher ratings of inattention and hyperactivity/impulsivity on the DBD and Attention Problems on the TRF in the prediction of consensus diagnosis. As seen at step 1 in Table 5, teacher ratings of inattention and hyperactivity/impulsivity on the DBD were each significantly associated with consensus diagnosis when entered simultaneously. However, as shown in step 2, in contrast to the model using parent ratings, ratings on the TRF-A did not significantly improve the logistic regression model (Table 5).

The third logistic regression model examined the contribution of parent reported symptoms on each ADHD dimension from the DISC (DISC-IA and DISC-HI) in the prediction of consensus diagnosis. As hypothesized, when entered simultaneously, each scale accounted for significant unique variance in the prediction of consensus diagnosis (Table 6).

Table 4

Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Parent Ratings on the Disruptive Behavior Disorders Rating Scale and the Child Behavior Checklist.

	Hosmer & Lemeshow χ^2	Nagelkerke R^2	Model χ^2	Block χ^2	β	SEβ	Wald	Odds Ratio	Percent Classified Correctly
Step 1	1.61	.90	272.66***	272.66***					95.1
PDBD-IA					3.22***	.96	11.20	24.94	
PDBD-HI					4.25***	1.12	14.37	69.87	
Step 2	2.51	.92	279.03***	6.37*					96.6
PDBD-IA					1.62	1.19	1.85	5.06	
PDBD-HI					4.33***	1.27	11.67	75.84	
CBCL-A					.17*	.08	4.83	1.18	

Note: β = logit coefficient. PDBD-IA: Parent Disruptive Behavior Disorders Rating Scale – Inattention Subscale. PDBD-HI: Parent Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. CBCL-A: Child Behavior Checklist – Attention Problems Syndrome Scale.

*ρ = α < .05; **ρ = α < .01; ***ρ = α < .001.

Table 5

Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis Teacher Ratings on the Disruptive Behavior Disorders Rating Scale and the Teacher Report Form.

	Hosmer & Lemeshow χ^2	Nagelkerke R^2	Model χ^2	Block χ^2	β	SEβ	Wald	Odds Ratio	Percent Classified Correctly
Step 1	7.46	.75	201.59***	201.59***					89.5
TDBD-IA					2.53***	.50	25.80	12.54	
TDBD–HI					1.32**	.45	8.51	3.75	
Step 2	8.26	.75	202.35***	.76					88.8
TDBD–IA					2.16***	.64	11.24	8.65	
TDBD-HI					1.10*	.52	4.50	3.01	
TRF-A					.07	.07	.74	1.07	

Note: β = logit coefficient. TDBD-IA: Teacher Disruptive Behavior Disorders Rating Scale – Inattention Subscale. TDBD-HI: Teacher Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. TRF-A: Teacher Report Form – Attention Problems Syndrome Scale.

*p = α < .05; **p = α < .01; ***p = α < .001.

Table 6

Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using the Parent Report on the Diagnostic Interview Schedule for Children-IV.

	Hosmer & Lemeshow χ^2	Nagelkerke R^2	Model χ^2	Block χ^2	β	SEβ	Wald	Odds Ratio	Percent Classified Correctly
Step 1	4.29	.88	260.27***	260.27***					94.0
DISC-IA					.56***	.13	18.23	1.75	
DISC-HI					1.10***	.22	24.12	3.00	

Note: β = logit coefficient. DISC-IA: Diagnostic Interview Schedule for Children-IV – Inattention Scale. DISC-HI: Diagnostic Interview Schedule for Children-IV –Hyperactivity/Impulsivity Scale.

*ρ = α < .05; **ρ = α < .01; ***ρ = α < .001.

Analyses Examining the Incremental Utility of Parent and Teacher Completed Rating Scales

The second set of logistic regressions examined the incremental utility of parent *and* teacher-completed rating scales in the prediction of consensus diagnosis. As an assessment of ADHD will almost uniformly include parent ratings of a child's behavior, we examined the incremental utility of teacher ratings beyond that of parent ratings (as we argue that parent ratings are the most efficient single method of assessment). Further, as symptom-based rating scales are more efficient than the longer, empirically-derived scales, they were entered into the model first. Specifically, we hypothesized that: 1) the addition of teacher-completed symptom-based rating scales, would contribute significantly to a model including parent-completed symptom-based scales and 2) that neither empirically-derived scale would contribute significant variance to a model containing both parent and teacher-completed symptom-based ratings.

Consistent with our model-building approach, at step 1, parent ratings of inattention and hyperactivity/impulsivity (PDBD-IA and PDBD-HI) were entered followed by teacher ratings of inattention and hyperactivity/impulsivity (TDBD-IA and TDBD-HI) at step 2. At step 3, we included ratings of Attention Problems on the CBCL and TRF. Consistent with our hypothesis, our analyses indicated that the addition of teacher ratings did result in statistically significant model improvement (Block $\chi^2 = 48.47$, $\rho < .001$). In fact, the logistic regression model including parent and teacher ratings of inattention and hyperactivity/impulsivity resulted in near-perfect model fit and association and correctly classified 265 of 267 participants. As would be expected given

these findings, the addition of the ratings on the CBCL-A and TRF-A did not significantly improve the overall model (Block $\chi^2 = .59$, $\rho = .756$) (Table 7). As discussed earlier, given the large logit coefficients within the model, insignificant Wald statistics were observed despite significant model and block chi-square values.

As a result of these issues and the uninterpretability of individual logit coefficients, we conducted a post-hoc conditional forward-entry logistic regression analysis including each of the parent and teacher DBD scales to examine their respective statistical contribution in the prediction of consensus diagnosis. Results indicated that parent ratings of inattention (PDBD-IA) entered the model first, followed by teacher ratings of hyperactivity/impulsivity (TDBD-HI) at step 2, and parent ratings of hyperactivity/impulsivity (PDBD-HI) at step 3 (Table 8). Teacher ratings of inattention (TDBD-IA) were not incrementally associated with consensus diagnosis. Subsequent post-hoc analyses examining the incremental utility of each of these three scales (PDBD-IA, PDBD-HI, and TDBD-HI), when added to a model including both other scales, indicated that the addition of teacher ratings of hyperactivity/impulsivity on the DBD resulted in a 4.1% increase in classification; whereas, parent ratings of inattention and hyperactivity/impulsivity on the DBD accounted for increases in classification of 1.5% and 0.4%, respectively.

Analyses Examining the Incremental Utility of Parent or Teacher Completed Rating Scales and a Structured Diagnostic Interview

The final set of logistic regression analyses examined the incremental utility of a

Table 7

Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Parent and Teacher Ratings on the Disruptive Behavior Disorders Rating Scales, the Child Behavior Checklist, and the Teacher Report Form.

	Hosmer & Lemeshow χ^2	Nagelkerke R^2	Model χ^2	Block χ^2	β	SEβ	Wald	Odds Ratio	Percent Classified Correctly
Step 1	1.61	.90	272.66***	272.66***					95.1
PDBD-IA					3.22***	.96	11.20	24.94	
PDBD-HI					4.25***	1.12	14.37	69.87	
Step 2	.62	.99	321.13***	48.47***					99.3
PDBD-IA					10.55	7.11	2.20	37995.47	
PDBD-HI					8.41	5.72	2.16	4474.60	
TDBD–IA					3.37	3.73	.82	29.05	
TDBD-HI					9.61	8.91	1.16	14906.83	
Step 3	.85	.99	321.72***	.59					99.3
PDBD-IA					18.60	16.83	1.22	125687.76	
PDBD-HI					11.82	10.72	1.22	135403.34	
TDBD–IA					3.34	5.60	.36	28.12	
TDBD-HI					10.94	8.67	1.59	56318.08	
CBCL-A					-.42	.60	.48	.66	
TRF-A					.20	.59	.11	1.22	

Note: β = logit coefficient. PDBD-IA: Parent Disruptive Behavior Disorders Rating Scale – Inattention Subscale. PDBD-HI: Parent Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. TDBD-IA: Teacher Disruptive Behavior Disorders Rating Scale – Inattention Subscale. TDBD-HI: Teacher Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. CBCL-A: Child Behavior Checklist – Attention Problems Syndrome Scale. TRF-A: Teacher Report Form – Attention Problems Syndrome Scale.

*ρ = α < .05; **ρ = α < .01; ***ρ = α < .001.

Table 8

Summary of Conditional Forward-Entry Logistic Regression Analysis Predicting Consensus Diagnosis using Parent and Teacher Ratings on the Disruptive Behavior Disorders Rating Scales.

	Hosmer & Lemeshow χ^2	Nagelkerke R^2	Model χ^2	Step χ^2	β	SEβ	Wald	Odds Ratio	Percent Classified Correctly
Step 1	8.02	.86	248.52***	248.52***					91.8
PDBD-IA					5.38***	.81	37.81	217.25	
Step 2	1.46	.97	310.64***	62.12***					98.9
PDBD-IA					10.89**	3.72	8.55	53587.57	
TDBD-HI					7.02**	2.61	7.21	112.71	
Step 3	.03	.98	319.45***	8.82**					99.3
PDBD-IA					12.41	7.99	2.41	244063.51	
TDBD-HI					8.29	5.12	2.63	3998.11	
PDBD-HI					13.98	9.87	2.01	1179050.60	

Note: β = logit coefficient. PDBD-IA: Parent Disruptive Behavior Disorders Rating Scale – Inattention Subscale. PDBD-HI: Parent Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. TDBD-HI: Teacher Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale.

*ρ = α < .05; **ρ = α < .01; ***ρ = α < .001.

structured diagnostic interview within a model including parent and/or teacher-completed rating scales. We hypothesized that the inattentive and hyperactive/impulsive symptoms endorsed on the structured diagnostic interview *would not* account for significant improvement within a model including parent symptom-based and empirically-derived ratings as parent-reported symptoms on the diagnostic structured interview would be redundant. Alternatively, we hypothesized that the inattentive and hyperactive/impulsive symptoms endorsed on the structured diagnostic interview *would* account for significant improvement when added to a model including teacher symptom-based and empirically-derived ratings given the non-redundancy of information by informant. Finally, given the robustness of the regression model including parent and teacher symptom-based ratings alone, we hypothesized that the addition of inattentive and hyperactive/impulsive symptoms endorsed on the DISC would not result in significant model improvement.

The first logistic regression analysis examined the incremental utility of parent-reported inattentive and hyperactive/impulsive symptoms endorsed on the DISC within a model including parent-completed symptom-based and empirically-derived rating scales. At step 1, parent ratings of inattention and hyperactivity/impulsivity on the DBD and ratings of Attention Problems on the CBCL were entered. At step 2, the number of endorsed symptoms on the inattention and hyperactivity/impulsivity scales of the DISC was entered (DISC-IA and DISC-HI). Results indicated that our hypothesis was not supported as the addition of the DISC-IA and DISC-HI scales did result in statistically significant improvement to the overall model (Block $\chi^2 = 9.85$; $\rho = .007$) (Table 9).

Table 9

Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Parent Ratings on the Disruptive Behavior Disorders Rating Scale, the Child Behavior Checklist, and Parent Report on the Diagnostic Interview Schedule for Children-IV.

	Hosmer & Lemeshow χ^2	Nagelkerke R^2	Model χ^2	Block χ^2	β	SEβ	Wald	Odds Ratio	Percent Classified Correctly
Step 1	2.51	.92	279.03***	279.03***					96.6
PDBD-IA					1.62	1.19	1.85	5.06	
PDBD-HI					4.33***	1.27	11.67	75.84	
CBCL-A					.17*	.08	4.83	1.18	
Step 2	.68	.93	288.81***	9.85**					96.3
PDBD-IA					2.17	1.67	1.69	8.79	
PDBD-HI					2.17	1.64	1.75	8.75	
CBCL-A					.14	.09	2.29	1.15	
DISC-IA					.05	.25	.04	1.05	
DISC-HI					.77*	.31	6.13	2.17	

Note: β = logit coefficient. PDBD-IA: Parent Disruptive Behavior Disorders Rating Scale – Inattention Subscale. PDBD-HI: Parent Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. CBCL-A: Child Behavior Checklist – Attention Problems Syndrome Scale. DISC-IA: Diagnostic Interview Schedule for Children-IV – Inattention Module. DISC-HI: Diagnostic Interview Schedule for Children-IV – Hyperactivity/Impulsivity Module.

*ρ = α < .05; **ρ = α < .01; ***ρ = α < .001.

A second logistic regression analysis was conducted to examine the incremental utility of the structured diagnostic interview beyond teacher-completed symptom-based and empirically-derived rating scales. Results were consistent with our hypotheses, as the inclusion of DISC-IA and DISC-HI scales resulted in significant model improvement when added to a model including teacher ratings of inattention and hyperactivity/ impulsivity on the DBD and Attention Problems on the TRF (Block $\chi^2 = 98.97$; $\rho < .001$) (Table 10).

The final logistic regression analysis examined the incremental utility of the structured diagnostic interview beyond that of parent and teacher-completed symptom-based rating scales. Specifically, at step 1, parent and teacher ratings of inattention and hyperactivity/impulsivity on the DBD (PDBD-IA, PDBD-HI, TDBD-IA, and TDBD-HI) were entered. At step 2, the number of endorsed symptoms on the inattention and hyperactivity/impulsivity scales of the DISC were entered (DISC-IA and DISC-HI). As hypothesized, the addition of the DISC-IA and DISC-HI scales did not result in significant model improvement (Block $\chi^2 = .91$, $\rho > .05$ (Table 11).

Examination of Logistic Regression Models including Significant Covariates

It should be noted that no covariates were included in the above logistical regression models despite significant group differences (namely, sex and a standardized measure of IQ). This decision was based on past reviews suggesting that the analysis of covariance is inappropriate when applied to situations involving non-random group assignments, as preexisting group differences cannot then be assumed to be independent

Table 10

Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Teacher Ratings on the Disruptive Behavior Disorders Rating Scale and Teacher Report Form and Parent Report on the Diagnostic Interview Schedule for Children-IV.

	Hosmer & Lemeshow χ^2	Nagelkerke R^2	Model χ^2	Block χ^2	β	SEβ	Wald	Odds Ratio	Percent Classified Correctly
Step 1	8.26	.75	202.35***	202.35***					88.8
TDBD–IA					2.16***	.64	11.24	8.65	
TDBD-HI					1.10*	.52	4.50	3.01	
TRF-A					.07	.07	.74	1.07	
Step 2	1.63	.96	301.32***	98.97***					97.0
TDBD–IA					4.39	1.86	5.53	80.20	
TDBD-HI					.39	1.13	.12	1.48	
TRF-A					-.04	.12	.12	.96	
DISC-IA					.35	.24	2.12	1.42	
DISC-HI					1.61	.48	11.50	5.00	

Note: β = logit coefficient. TDBD-IA: Teacher Disruptive Behavior Disorders Rating Scale – Inattention Subscale. TDBD-HI: Teacher Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. TRF-A: Teacher Report Form – Attention Problems Syndrome Scale. DISC-IA: Diagnostic Interview Schedule for Children-IV – Inattention Module. DISC-HI: Diagnostic Interview Schedule for Children-IV – Hyperactivity/Impulsivity Module.

*$\rho = \alpha < .05$; **$\rho = \alpha < .01$; ***$\rho = \alpha < .001$.

Table 11

Summary of Hierarchical Logistic Regression Analysis Predicting Consensus Diagnosis using Parent and Teacher Ratings on the Disruptive Behavior Disorders Rating Scales and Parent Report on the Diagnostic Interview Schedule for Children-IV.

	Hosmer & Lemeshow χ^2	Nagelkerke R^2	Model χ^2	Block χ^2	β	SEβ	Wald	Odds Ratio	Percent Classified Correctly
Step 1	.62	.99	321.13***	321.13***					99.3
PDBD-IA					10.55	7.11	2.20	37995.47	
PDBD-HI					8.41	5.72	2.16	4474.60	
TDBD-IA					3.37	3.73	.82	29.05	
TDBD-HI					9.61	8.91	1.16	14906.83	
Step 2	1.63	.99	322.03***	.91					98.9
PDBD-IA					14.28	10.41	1.88	1598211.60	
PDBD-HI					8.52	11.43	.56	5029.29	
TDBD-IA					5.23	4.98	1.11	186.99	
TDBD-HI					9.37	10.07	.87	11741.69	
DISC-IA					-1.01	1.21	.69	.37	
DISC-HI					.39	1.87	.04	1.48	

Note: β = logit coefficient. PDBD-IA: Parent Disruptive Behavior Disorders Rating Scale – Inattention Subscale. PDBD-HI: Parent Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. TDBD-IA: Teacher Disruptive Behavior Disorders Rating Scale – Inattention Subscale. TDBD-HI: Teacher Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. DISC-IA: Diagnostic Interview Schedule for Children-IV – Inattention Module. DISC-HI: Diagnostic Interview Schedule for Children-IV –Hyperactivity/Impulsivity Module.

*ρ = α < .05; **ρ = α < .01; ***ρ = α < .001.

of the predictor variables (Miller and Chapman, 2001). For the skeptical reader, however, all regression models also were conducted including sex and IQ as covariates. These models did not differ markedly from the above regression models and as such are not reported here.

Examination of the Sensitivity, Specificity, Classification Rates, and Odds Ratios within Logistic Regression Models.

Within each regression, the chance-corrected rate of agreement (kappa); the sensitivity and specificity of each model; and estimated odds ratios were calculated. Following our previous discussion regarding models demonstrating large effects and resulting in insignificant Wald statistics and beta coefficients, the odds ratios within models exhibiting these problems are not discussed. For example, logistic regression models utilizing both parent *and* teacher ratings on the DBD resulted in insignificant logit coefficients and extremely high odds ratios (Table 7). As such, these odds ratios are not discussed due to their lack of interpretability. However, this model also demonstrated the highest rates of chance-corrected agreement ($\kappa = .98$), sensitivity/specificity (.99 and .99, respectively), and overall model significance.

As seen in Table 12, examination of the primary logistic regression models indicated satisfactory rates of agreement, sensitivity, and specificity across all methods (Table 12). In particular, models including parent ratings of inattention and hyperactivity/impulsivity consistently demonstrated the most robust rates of agreement, sensitivity, and specificity. Results were consistent, albeit slightly less robust, when examining teacher ratings of inattention and hyperactivity/impulsivity on the DBD. The

Table 12

Summary of Sensitivity, Specificity, Percent Agreement, and Chance-Corrected Agreement for Logistic Regression Models using Parent and Teacher Ratings on the Disruptive Behavior Disorders Rating Scales, Child Behavior Checklist, and Teacher Report Form, and Parent Report on the Diagnostic Interview Schedule for Children-IV.

	Sensitivity	Specificity	Percent Classified Correctly	Kappa
PDBD-IA and PDBD-HI	.97	.92	95%	.89
PDBD-IA, PDBD-HI, and CBCL-A	.98	.94	97%	.92
TDBD-IA and TDBD-HI	.92	.83	90%	.78
TDBD-IA, TDBD-HI, and TRF-A	.91	.83	89%	.74
DISC-IA and DISC-HI	.96	.90	94%	.86
PDBD-IA, PDBD-HI, CBCL-A, DISC-IA and DISC-HI	.97	.94	96%	.91
TDBD-IA, TDBD-HI, TRF-A, DISC-IA, and DISC-HI	.98	.96	97%	.94
PDBD-IA, PDBD-HI, TDBD-IA, and TDBD-HI	.99	.99	99%	.98

Note: PDBD-IA: Parent Disruptive Behavior Disorders Rating Scale – Inattention Subscale. PDBD-HI: Parent Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. TDBD-IA: Teacher Disruptive Behavior Disorders Rating Scale – Inattention Subscale. TDBD-HI: Teacher Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. CBCL-A: Child Behavior Checklist – Attention Problems Syndrome Scale. TRF-A: Teacher Report Form – Attention Problems Syndrome Scale. DISC-IA: Diagnostic Interview Schedule for Children-IV – Inattention Scale. DISC-HI: Diagnostic Interview Schedule for Children-IV –Hyperactivity/Impulsivity Scale.

addition of empirically-derived rating scales to these models did not consistently improve rates of agreement, sensitivity, or specificity. Finally, examination of the DISC scales of inattention and hyperactivity/impulsivity also indicated robust rates of agreement, sensitivity, and specificity. Further, the addition of the DISC ratings within models including behavioral rating scales did contribute to increased, albeit minimal, rates of agreement, sensitivity, and specificity.

Examination of the Validity and Clinical Utility of Diagnostic Algorithms

The final series of analyses explored the clinical utility of diagnostic algorithms employing the measures that demonstrated the greatest association with consensus diagnosis in the above logistic regression models. We hypothesized that employing diagnostic algorithms would provide equivalent rates of diagnostic classification while increasing the clinical utility of each scale, as utilizing logistic regression models as a means of diagnostic classification is untenable in a clinic setting. As previous research has primarily examined the utility of inattentive or hyperactive/impulsive symptoms separately (Powers et al., 1998; Powers, et al. 2001), we examined diagnostic algorithms that utilized ratings of both inattentive and hyperactive/impulsive symptoms. Algorithms were examined in order of efficiency (i.e., most efficient to least efficient), until classification and sensitivity/specificity rates were equivalent to the above logistic regression models.

Diagnostic algorithms were derived following DSM-IV criteria by requiring a minimum of six symptoms on each of the ADHD dimensions to be endorsed as present for a child to be classified as ADHD. Following the findings of Power and colleagues

(Power et al., 1998a; Power et al., 1998b; Power, et al., 2001) and general consensus (Pelham et al., 2005), symptoms endorsed as "Pretty Much" or "Very Much" on the DBD scales were considered present. We also required that these "stringent" algorithms derive classifications parallel with the consensus diagnoses (e.g., endorsement of at least six symptoms in either hyperactivity/ impulsivity or *both* inattention *and* hyperactivity/impulsivity). However, as behavioral rating scales are frequently used as screening measures for ADHD, we also examined the rate of agreement between consensus diagnosis and *any* algorithm-derived diagnosis of ADHD (e.g., ADHD-IA, ADHD-HI, *or* ADHD-C). Consequently, a second "lenient" algorithm was used that only required the endorsement of six symptoms of inattention *or* hyperactivity/impulsivity to receive a classification as ADHD. Both the "stringent" and "lenient" algorithm-derived classifications were then compared with consensus diagnosis.

Each algorithm utilized the most efficient methods that demonstrated incremental utility in the logistic regression models. We then proceeded to continue requiring additional (and less efficient) measures in each subsequent algorithm. Thus, we continued deriving less-efficient, but increasingly comprehensive algorithms until maximal agreement between the consensus and algorithm-derived diagnoses was met. As seen in Table 13, as the lenient diagnostic algorithm employing all behavioral rating scales resulted in perfect agreement, no further diagnostic algorithms were examined. However, all preceding algorithms are discussed below.

The initial algorithms employed the parent ratings of inattention and hyperactivity/impulsivity on the DBD to derive diagnostic classifications following the

Table 13

Summary of Sensitivity, Specificity, Percent Agreement, and Chance-Corrected Agreement for Diagnostic Algorithms utilizing Parent and Teacher Ratings on the Disruptive Behavior Disorders Rating Scales, the Child Behavior Checklist, and the Teacher Report Form.

	Sensitivity	Specificity	Percent Classified Correctly	Kappa
Stringent Diagnostic Algorithm Utilizing PDBD-IA and PDBD-HI	.46	1.00	63%	.45
Lenient Diagnostic Algorithm Utilizing PDBD-IA and PDBD-HI	.81	1.00	87%	.74
Stringent Diagnostic Algorithm Utilizing TDBD-IA and TDBD-HI	.41	1.00	59%	.42
Lenient Diagnostic Algorithm Utilizing TDBD-IA and TDBD-HI	.72	.99	80%	.65
Stringent Diagnostic Algorithm Utilizing PDBD-IA, PDBD-HI, TDBD-IA and TDBD-HI	.96	.92	95%	.89
Lenient Diagnostic Algorithm Utilizing PDBD-IA, PDBD-HI, TDBD-IA and TDBD-HI	.98	.92	96%	.92
Stringent Diagnostic Algorithm Utilizing PDBD-IA, PDBD-HI, TDBD-IA, TDBD-HI, CBCL-A, and TRF-A	.91	1.00	94%	.88
Lenient Diagnostic Algorithm Utilizing PDBD-IA, PDBD-HI, TDBD-IA, TDBD-HI, CBCL-A, and TRF-A	1.00	1.00	100%	1.00

Note: PDBD-IA: Parent Disruptive Behavior Disorders Rating Scale – Inattention Subscale. PDBD-HI: Parent Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. TDBD-IA: Teacher Disruptive Behavior Disorders Rating Scale – Inattention Subscale. TDBD-HI: Teacher Disruptive Behavior Disorders Rating Scale – Hyperactivity/Impulsivity Subscale. CBCL-A: Child Behavior Checklist – Attention Problems Syndrome Scale. TRF-A: Teacher Report Form – Attention Problems Syndrome Scale.

criteria described above. Utilizing parent ratings alone, the stringent diagnostic algorithm correctly classified 62.5% of all participants. In contrast, utilizing the lenient diagnostic algorithm resulted in the correct classification of 86.5% of participants. Of note, utilizing either the stringent or lenient algorithm resulted in zero "false positive" diagnoses (e.g., participants without consensus diagnosis of ADHD placed in ADHD group).

These procedures were then repeated with teacher ratings of inattention and hyperactivity/impulsivity on the DBD. Findings, although consistent, were not as robust as when parent ratings were used in classifying participants. Specifically, 58.8% of cases were classified correctly utilizing the stringent diagnostic algorithm; whereas, concordance rates increased substantially (80.1%) when the lenient diagnostic algorithm was employed. We then utilized a diagnostic algorithm requiring both parent *and* teacher ratings on the DBD. As no clear guidelines are set forth by the DSM-IV in combining information from multiple informants in the assessment of ADHD, we utilized a flexible algorithmic approach similar to that used in previous studies of ADHD (Rowland, Umbach, Catoe, et al., 2001; Wolraich, et al., 2004;). Specifically, to ensure the presence of symptoms in multiple settings, we required a minimum of six inattentive and six hyperactivity/impulsivity symptoms to be endorsed as present by one informant with at least three inattention and three hyperactivity/impulsivity symptoms be endorsed by the other informant. For example, if six symptoms of inattention and three symptoms of hyperactivity/impulsivity were endorsed by a child's parent and three symptoms of inattention and six symptoms of hyperactivity/impulsivity were endorsed by the child's teacher, then the child was classified as ADHD. Utilizing this stringent diagnostic

algorithm resulted in correct classification of 94.8% of participants (sensitivity = .96; specificity = .92). However, utilizing a lenient diagnostic algorithm requiring only the endorsement of inattentive *or* hyperactive/impulsive symptoms following the above criteria resulted in correct classification of 96.2% of participants.

Finally, we examined the clinical utility of an algorithm that required information from multiple informants and utilized both symptom-based and empirically-derived rating scales. In addition to the criteria described for the previous algorithm, we now required a T-score greater than or equal to 70 on either the CBCL-A or TRF-A for a child to be classified as ADHD. (Stringent and lenient diagnostic procedures were identical to the previous algorithm with the addition of the CBCL/TRF requirement). Our findings indicated that 94.4% of participants were classified correctly utilizing the stringent diagnostic algorithm. However, utilizing the lenient criterion resulted in all 267 participants being classified correctly (sensitivity = 1.00; specificity = 1.00). As this lenient algorithm resulted in perfect agreement with the consensus diagnosticians, no further algorithms were examined.

Discussion

The primary goals of the present study were threefold: 1) to examine the *independent* predictive utility of symptom-based rating scales, empirically-derived rating scales, and a structured diagnostic interview in a multi-method, multi-informant assessment of ADHD; 2) to examine the *incremental* predictive utility of each method and informant in the prediction of diagnostic status (e.g., ADHD or Not ADHD); and 3) to examine the incremental utility and efficiency of clinically-relevant algorithms using the methods that demonstrated the greatest statistical association with a diagnosis of ADHD.

Consistent with our expectations and previous research (Crystal, Ostrander, Chen, & August, 2001; Pelham et al., 2005;), our results indicated when methods were examined incrementally, results consistently supported the superiority of both parent-completed methods and of symptom-based rating scales. Results were less consistent regarding the incremental utility of empirically-derived rating scales beyond symptom-based ratings. Specifically, the Attention Problems Syndrome Scale on the CBCL was related to improved model fit; whereas, ratings on the Attention Problems Syndrome Scale on the TRF did not result in significant model improvement. As hypothesized, a structured diagnostic interview did not significantly improve a regression model containing symptom-based rating scales completed by a child's parent *and* teacher. In fact, utilizing both parent and teacher ratings of inattention and hyperactivity/impulsivity on a symptom-based scale resulted in the correct classification of all but two cases in our sample. As such, information provided by other methods provided only redundant

information in the prediction of a diagnosis of ADHD (as virtually all symptom information necessary for diagnosis was provided by parent and teacher symptom-based ratings). This finding provides empirical support for Pelham's assertion that a structured diagnostic interview accounts for little incremental utility beyond that accounted for by the more efficient parent and teacher symptom-based rating scales in the assessment of ADHD (Pelham et al., 2005). However, the structured diagnostic interview did account for significant improvement in logistic regression models containing either parent *or* teacher ratings on symptom-based and empirically-derived scales. However, as logistic regression models utilizing symptom-based ratings were particularly robust in predicting diagnosis, information provided by a structured diagnostic interview, although *statistically* significant, accounted for minimal increases in *actual* diagnostic classification when added to a model including teacher ratings (21 participants) or a model including parent ratings (0 participants).

To our knowledge, this is the first study to examine the incremental utility of a structured diagnostic interview *within* a comprehensive assessment of ADHD. Following Pelham and colleagues' assertion that "diagnosing ADHD is most efficiently accomplished with parent and teacher rating scales" (Pelham et al., 2005, p. 469), our findings indicated that utilizing only three nine-item scales across two informants led to the correct classification of 265 of our 267 participants. As such, the incremental utility of a structured diagnostic interview (or any assessment method for that matter) was theoretically futile in relation to ADHD symptomatology. Further, parent ratings on a

symptom-based rating scale alone resulted in better fit, greater association, and greater classification than a structured diagnostic interview alone within our sample.

In the context of this result, we raise two issues for discussion and possible consideration in future research. First, the DISC interview was completed at the end of a comprehensive assessment that included multiple behavioral rating scales and a semi-structured clinical interview reviewing primary concerns including onset, frequency and intensity of problem behaviors. Further, consistent with the default ordering of modules in the DISC-IV, the ADHD modules were presented at the end of the structured interview (after approximately 85% of the interview had been completed). Given past findings regarding symptom attenuation within structured interviews (Jensen & Edelbrock, 1999; Piacentini, et al., 1999), it is possible that the frequency of symptom endorsement declined over the course of the structured interview in this study. Future studies might consider administration of the ADHD module first to eliminate this concern. Second, given the criteria necessary for a symptom to be noted as present on the DISC-IV (present across settings for at least six months), it is possible that some symptoms rated as "Pretty Much" a problem on the DBD rating scales may not have been endorsed on the DISC-IV. Given the superiority of parent-reported symptoms on the DBD versus parent report of symptoms on the DISC in our study, the assumption of equivalence between ratings on symptom-based rating scales and their relative endorsement on a structured diagnostic interview should be examined further.

Based on our findings, this study provides continued support regarding the validity and efficiency of behavioral rating scales in the assessment of ADHD (Pelham,

2005; Power et al., 2001). Our results indicated that utilizing either rater (parent or teacher) on a scale of inattentive or hyperactive/impulsive symptoms resulted in rates of agreement ranging from 87.3% to 93.6% with diagnoses derived from a comprehensive, "gold standard assessment of ADHD. Also, consistent with previous literature, utilizing multiple raters significantly increased diagnostic efficiency and classification rates (Power et al., 1998a; Power et al., 1998b). However, closer examination revealed that these findings were driven exclusively by parent ratings of inattentive and hyperactive/ impulsive symptoms and teacher ratings of hyperactive/impulsive symptoms of ADHD (see Table 8). Utilizing these three scales alone resulted in maximal rates of agreement with consensus diagnoses of ADHD-C and ADHD-HI and accounted for virtually all of the variance within the logistic regression model. This finding is somewhat surprising given previous findings suggesting the superiority of teacher ratings of inattention in predicting ADHD diagnosis (as teachers place more demands on children for sustained attention in school than parents at home) (Landau, Lorch, & Milich, 1992). However, we argue that this finding is partially attributable to our sample as only children with ADHD-Combined Type or ADHD-Hyperactive/Impulsive Type were included within the study; past research supports the superiority of symptoms of hyperactivity/impulsivity in predicting a diagnosis of ADHD in similar samples (Simonsen & Bullis, 2007). Further, teacher ratings of inattention were independently associated with ADHD diagnosis and contributed to significantly greater fit in regression models excluding any one of the three scales noted above. Importantly, no other combination of measures, methods, or raters was able to correctly classify more participants, result in greater fit, or account for greater

variance than symptom-based rating scales. Surprisingly, this outcome remained true even when combining the ratings on a parent-informant structured diagnostic interview and teacher-completed rating scales. This finding suggests that the information provided by symptom-based rating scales not only uniquely informs diagnosis, but is essential in the assessment of ADHD and likely renders the symptom information provided by a structured diagnostic interview redundant (Pelham et al., 2005).

Our findings also provide continued support for the use of multiple informants in the assessment of ADHD. Following DSM-IV guidelines and near universal agreement affirming the collection of information from multiple informants in multiple settings, the importance of teacher ratings in diagnostic classification was supported in this study. In fact, despite the notable robustness of models including parent ratings alone, the addition of teacher ratings contributed significantly to greater model fit. However, it should be noted that parent ratings were more highly associated with diagnosis and led to greater rates of classification than teacher ratings alone. Although only teacher ratings of hyperactivity/impulsivity significantly improved a model already including parent ratings, we argue for the continued inclusion of teacher ratings of inattention given their independent association with an ADHD diagnosis in this study as well as the litany of research documenting their predictive utility (Power, et al., 2001; Power, et al., 1998a). Further, as our sample included only children with ADHD-C or ADHD-HI, we argue that the inclusion of teacher ratings of inattention may be even more critical in assessing for inattentive only type ADHD.

Finally, this study demonstrated both the statistical validity *and* clinical utility of behavioral rating scales in the assessment of ADHD. Our examination of clinically-relevant diagnostic algorithms provided strong empirical support for their use as a proxy for the robust logistic regression models. In fact, our data indicated that a diagnostic algorithm utilizing behavioral rating scales alone was able to correctly classify all 267 of our participants correctly. This finding strongly supports previous research suggesting that little to no additional information is added to the prediction of an ADHD diagnosis beyond that provided by behavioral rating scales (Pelham, et al., 2005; Power, et al., 2001; Simonsen & Bullis, 2007). However, our findings pointed out the necessity of CBCL or TRF ratings in the diagnostic algorithm for correct classification of all 267 participants. In contrast, algorithms utilizing only parent and teacher symptom-based rating scales correctly classified 254 of the 267 participants.

We argue that these findings provide strong support for the value of diagnostic algorithms that rely on valid, cost-efficient methods in the assessment of ADHD. Further, although it is theoretically feasible for clinicians to utilize regression models to inform diagnostic decision-making, in reality this process is likely untenable. We contend that this examination of diagnostic algorithms moves incremental validity research into the realm of clinical relevance through considering cost-efficiency, practicality, and real-world utility (Johnston & Murray, 2003). Specifically, as suggested by previous research (Power et al., 2001; Sayal, Letch, & Abd, 2008; Simonsen & Bullis, 2007), our findings provide support for the use of a multiple-gating procedure in the assessment of ADHD. Given the robustness of rating scales in our sample and their

associated efficiency, strategies that utilize optimal cut points on rating scales that maximize sensitivity while maintaining acceptable levels of specificity would likely contribute to increasing the efficiency of an ADHD assessment. Utilizing a multiple-gate procedure would likely decrease the expenditure of unnecessary resources required for "gold standard" assessments (especially assessments requiring a structured diagnostic interview). For example, children likely to meet criteria for ADHD may require only parent and teacher rating scale data and a semi-structured clinical interview (assessing developmental, social, academic, and family functioning as well as age of onset and impairment). This approach would allow for other resources to be applied towards treatment planning and intervention (Pelham, et al., 2005). Similarly, children unlikely to meet criteria for a diagnosis of ADHD may be identified using efficient behavioral rating scales without requiring further parent, teacher, child, or clinician resources. As the purpose of this study was not to identify optimal cut points, this question should be examined further. However, we argue that behavioral rating scales do provide the most efficient procedure for assessing children with ADHD.

Strengths and Limitations

This study had several notable strengths and limitations. First, determinations about group membership (e.g., diagnostic status) were based on a consensus decision agreed upon by two psychologists after reviewing information provided by a comprehensive evaluation strategy that included parent and teacher rating scales, a structured diagnostic interview, a full cognitive and achievement battery, child self-report measures, and a clinical interview with the child's parent assessing his/her developmental,

social, and academic functioning. Despite the presumed validity of this "gold standard" diagnostic procedure, the results may have differed if an alternative diagnostic procedure was used as the criterion for an accurate diagnosis. Although we view the comprehensive diagnostic procedure used as a strength of our study, one may argue that the use of a set of predictors that also were utilized to derive the criterion may contribute to the high rates of agreement found. However, in the absence of a known objective alternative procedure or "test" for establishing with certainty the presence of an ADHD diagnosis, we cannot suggest an alternative that might have been better.

Although this study included a significant percentage of females (22.5%) and children of differing races/ethnicities (80.5% Caucasian; 9.7% African-American; 9.8% Other), there were neither enough girls nor enough minority children to permit comparative analysis of the incremental utility of each method for these subsets of the sample. Future research should examine whether the incremental utility of rating scales and the usefulness of a diagnostic algorithm in the assessment of ADHD is consistent across sex and racial/ethnic backgrounds.

Clinical and Research Implications

In summary, the goals of this study were to examine the actual, unique contributions of universally recommended assessment methods in a comprehensive, "gold standard" assessment of ADHD. In relation to these goals, this study demonstrated the independent contributions of behavioral rating scales and a structured diagnostic interview in the assessment of ADHD. Further, we demonstrated the relative incremental utility of these methods across informants in the prediction of a diagnosis of ADHD. As

such, this research challenged the practice of using a structured diagnostic interview, in addition to behavioral rating scales, as an efficient and incrementally valid method of assessment. In fact, the optimal methods necessary to inform an assessment of ADHD in this study were parent and teacher symptom-based rating scales alone. These results display the value of behavioral rating scales and the relative limitations of much less efficient structured diagnostic interviews. As an understanding of the incremental validity of assessment methodology is critical in bridging the gap between laboratory and clinic-based settings, we argue that future research should incorporate similar approaches to identify the most ecologically valid, cost-effective methods in assessment.

Further, this study provided a clinically relevant, efficient method of integrating information from behavioral rating scales completed by multiple informants in the assessment of ADHD. Importantly, a relatively simple diagnostic algorithm utilizing symptom-based and empirically-derived scales was able to classify *all* participants within our sample correctly. Although this finding may not be as robust in settings where complex comorbidities are more common, future researchers are strongly encouraged to examine strategies for integrating assessment methods to facilitate the development of valid, efficient diagnostic algorithmic that identify ADHD efficiently in a manner that is easily generalizable to a clinic-based setting.

Although the results of this study support the practice of requiring multiple informants' reports of ADHD symptoms, the robust predictive utility of parent ratings alone suggests their possible use within a multiple-gating procedure maximizing efficiency. This argument is consistent with previous research arguing for a single-

informant to either rule-out a diagnosis of ADHD or inform the need for a more intensive assessment (Pelham, et al., 2005; Power et al., 2001; Simonsen & Bullis, 2001). However, as this study was not designed to identify optimal cut points, additional research verifying the use of such strategies and their respective clinical utility (e.g., incremental *and* ecological validity) in ruling in or out a diagnosis of ADHD is needed.

REFERENCES

Achenbach, T.M., & Rescorla, L.A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Agresti, A. (1996). *An introduction to categorical data analysis*. NY: John Wiley.

Albert, A., & Anderson, J.A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika, 71*, 1-10.

Allison, P. (2008). *Convergence failures in logistic regression*. Paper presented at the SAS Global Forum, San Antonio, TX.

American Academy of Pediatrics. (2000). Clinical practice guideline: Diagnosis and evaluation of the child with attention deficit/hyperactivity disorder. *Pediatrics, 105*, 1158-1170.

American Academy of Pediatrics. (2001). Clinical practice guideline: Treatment of the school-aged child with attention deficit hyperactivity disorder. *Pediatrics, 108*, 1033-1044.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Angold, A., Costello, E.J., Farmer, E.M.Z., Burns, B.J., & Erkanli, A. (1999). Impaired but undiagnosed. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 129-137.

Atkins, M.S., Pelham, W.E., & Licht, M.H. (1985). A comparison of objective classroom measures and teacher ratings of attention deficit disorder. *Journal of Abnormal Child Psychology, 13*, 155-167.

Barkley, R.A. (1988). Attention-deficit hyperactivity disorder. In E.J. Mash & L.G. Terdal (Eds.), *Behavioral assessment of childhood disorders* (2nd ed., pp. 113-155). New York: Guilford.

Barkley, R.A. (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *Journal of Abnormal Child Psychology, 19*, 149-178.

Barkley, R.A., & Grodzinksy, G.M. (1994). Are tests of frontal lobe functions useful in the diagnosis of attention deficit disorders? *Clinical Neuropsychologist, 8*, 121-139.

Barkley, R.A., Fischer, M., Edelbrock C., & Smallish, L. (1990). The adolescent outcome of hyperactive children diagnosed by research criteria, I: an 8-year prospective follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*, 546 – 557.

Barkley, R.A., Fischer, M., Smallish, L., & Fletcher, K. (2004). Young adult follow-up of hyperactive children: antisocial activities and drug use. *Journal of Child Psychology and Psychiatry, 45*, 195-211.

Barkley, R.A., Shelton, T.L., Crosswait, C., Moorehouse, M., Fletcher, K., Barrett, S.,

Jenkins, L., & Metevia, L. (2000). Multi-method psycho-educational intervention for preschool children with disruptive behavior: Preliminary results at post-treatment. *Journal of Child Psychology and Psychiatry, 41*, 319-332.

Crystal, D., Ostrander, R., Chen, R., & August, G. (2001). Multimethod assessment of psychopathology among DSM-IV subtypes of chilren with Attention-Deficit/Hyperactivity Disoder: self, parent, and teacher reports. *Journal of Abnormal Child Psychology, 29*, 189-205.

Chen, W.J., Faraone, S.V., Biederman, J., & Tsuang, M.T. (1994). Diagnostic accuracy of the Child behavior Checklist scales for Attention-Deficit Hyperactivity Disorder: A receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology, 62*, 1017-1025.

Cohen, P., O'Conner, P., Lewis, S., Velez, C.N., & Malachowski, B. (1987). Comparison of the DISC and K-SADS-P interviews of an epidemiological sample of children. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 662-667.

Collett, B.R., Ohan, J.L., & Myers, K.M. (2003). Ten-year review of rating scales. V: Scales assessing Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 1015-1037.

Costello, E.J., Edelbrock, C.S., & Costello, A.J. (1985). Validity of the NIMH Diagnostic Interview Schedule for Children-IV: A comparison between psychiatric and pediatric referrals. *Journal of Abnormal Child Psychology, 13*, 579-595.

DuPaul, G.J. (1991). Parent and teacher ratings of ADHD symptoms: Psychometric properties in a community-based sample. *Journal of Clinical Child Psychology, 20*, 245-253.

DuPaul, G.J., Power, T., Anastopoulos, A.D., & Reid, R. (1998). *Manual for the ADHD Rating Scale-IV.* New York: Guilford Press.

DuPaul, G.J., Power., T.J., McGoey, K.E., Ikeda, M.J., & Anastopoulos, A.D. (1998). Reliability and validity of parent and teacher ratings of Attention-Deficit/Hyperactivity Disorder symptoms. *Journal of Psychoeducational Assessment, 16*, 55-68.

Eiraldi, R.B., Power, T.J., Karustis, J.L., & Goldstein, S.G. (2000). Assessing ADHD and comorbid disorders in children: The Child Behavior Checklist and the Devereux Scales of Mental Disorders. *Journal of Clinical Child Psychology, 29*, 3-16.

Halperin, J., Sharma, V., Greenblatt, E., & Schwartz, S. (1991). Assessment of the continuous performance test: reliability and validity in a non-referred sample. *Journal of Consulting and Clinical Psychology, 3*, 601-608.

Hartman, R.R., Stage, S.A., & Webster-Stratton, C. (2003). A growth curve analysis of parent training outcomes: Examining the influence of child risk factors (inattention, impulsivity, and hyperactivity problems), parental, and family risk factors. *Journal of Child Psychology and Psychiatry, 44*, 388-398.

Heinz, G, & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine, 30*, 2409-2419.

Hintze, J.M., & Matthews, W.J. (2004). The generalizability of systematic direct observations across time and setting: A preliminary investigation of the psychometrics of behavioral observation. *School Psychology Review, 33*, 258-270.

Hodges, K. (1993). Structured interviews for assessing children. *Journal of Child Psychology and Psychiatry, 34*, 49-68.

Hosmer, D., & Lemeshow, S. (2000). *Applied Logistic Regression*, 2nd ed. NY: Wiley and Sons.

Jensen, P., & Edelbrock, C. (1999). Subject and interview characteristics affecting reliability of the Diagnostic Interview Schedule for Children-IV. *Journal of Abnormal Child Psychology, 27*, 413-415.

Jensen, P.S., Koretz, K., Locke, B., Schneider, S., Radke-Yarrow, M., Richters, J., et al. (1993). Child and adolescent psychopathology research: Problems and prospects for the 1990's. *Journal of Abnormal Child Psychology, 21*, 551-580.

Jensen, P.S., Wantanabe, J., Richters, J., Roper, N., Hibbs, E., Salzberg, A., et al. (1996). Scales, diagnoses, and child psychopathology: II. Comparing the CBCL and the DISC against external validators. *Journal of Abnormal Child Psychology, 24*, 151-168.

Johnston, C., & Murray, C. (2003). Incremental validity in the psychological assessment of children and adolescents. *Psychological Assessment, 15*, 496-507.

Landau, S., Lorch, E.P., & Milich, R. (1992). Visual attention to and comprehension of

television in attention-deficit hyperactivity disordered and normal boys. *Child Development, 63*, 928-937.

Loeber, R., Green, S.M., Lahey, B.B., Stouthamer-Loeber, M. (1991). Differences and similarities between children, mothers, and teacher as informants on disruptive child behavior. *Journal of Abnormal Child Psychology, 19*, 75-95.

Mannuzza, S., & Klein, R.G. (1999). Adolescent and adult outcomes in Attention-Deficit/Hyperactivity Disorder. In H.C. Quay & A.E. Hogan (Eds.). *Handbook of Disruptive Behavior Disorders* (pp. 279-294). New York: Kluwer Academic.

Mash, E.J., & Terdal, L.G. (1997). Assessment of child and family disturbance: A behavioral-systems approach. In E.J. Mash & L.G. Terdal (Eds.), *Assessment of Childhood Disorders*, 3rd ed. (pp. 3-68). New York: Guilford Press.

Massetti, G.M., Pelham, W.E., Chacko, A., Walker, K.S., Arnold, F.W., Coles, E.K., et al. (2003, November). *Situational variability of ADHD, ODD, and CD: Psychometric properties of the DBD interview and rating scale*. Poster presented at the Association for Advancement of Behavior Therapy Conference, Boston.

Mather, N., & Woodcock, R.W. (2001). Examiner's manual. *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.

McClellan, J.M., & Werry, J.S. (2000). Introduction: Research psychiatric diagnostic interviews for children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 19-27.

McConaughy, S.H. (2000). Self-reports: Theory and practice in interviewing children.

In E.S. Shapiro & T.R. Kratochwill (Eds.), *Behavioral Assessment in Schools: Theory, Research, and Clinical Foundations* (pp. 323 -351). New York: Guilford Press.

McConaughy, S. & Achenbach, T. (2004). *Manual for the ASEBA Test Observation Form.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

McGough, J.J., & McCracken, J.T. (2000). Assessment of Attention Deficit Hyperactivity Disorder: A review of recent literature. *Current Opinion in Pediatrics, 12*, 319-324.

Menard, S. (2002). *Applied logistic regression analysis, 2nd Ed.* Thousand Oaks, CA: Sage Publications.

Milich, R., Balentine, A.C., & Lynam, D.R. (2001). ADHD Combined Type and ADHD Predominantly Inattentive Type are distinct and unrelated disorders. *Clinical Psychology: Science and Practice, 8*, 463 – 488.

Miller, G.A. & Chapman, J.P. (2001). Misunderstanding Analysis of Covariance. *Journal of Abnormal Psychology, 110*, 40-48.

MTA Cooperative Group. (1999). 14-month randomized clinical trial of treatment strategies for children with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 56*, 1073-1086.

Ostrander, R., Weinfurt, K.P., Yarnold, P.R., & August, G.J. (1998). Diagnosing attention deficit disorders with the Behavioral Assessment System for Children and the Child Behavior Checklist: Test and construct validity analyses using

optimal discriminant classification trees. *Journal of Consulting and Clinical Psychology, 66*, 660-672.

Owens, J.S., & Hoza, B. (2003). The role of inattention and hyperactivity/impulsivity in the positive illusory bias. *Journal of Consulting and Clinical Psychology, 71*, 680-691.

Pelham, W.E., & Evans, S. (1992). Teacher ratings of DSM-III-R symptoms for the Disruptive Behavior Disorderss: prevalence, factor analyses, and conditional probabilities in a special education sample. *School Psychology Review, 21*, 285-300.

Pelham, W.E., Fabiano, G.A., & Massetti, G.M. (2005). Evidence-based assessment of Attention Deficit Hyperactivity Disorder in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*, 449-476.

Pelham, W. E., Gnagy, E. M., Greenslade, K. E. Milich, R. (1992). Teacher ratings of DSM-III—R symptoms for the Disruptive Behavior Disorderss. *Journal of the American Academy of Child & Adolescent Psychiatry, 31*, 210-218.

Piacentini, J., Roper, M., Jensen, P., Lucas, C., Fisher, P., Bird, H., et al. (1999). Informant-based determinants of symptom attenuation in structured child psychiatric interviews. *Journal of Abnormal Child Psychology, 27*, 417-428.

Piacentini, J., Shagger, D., Fisher, P., Schwab-Stone, M. Davies, M., & Gioia, P. (1993). The Diagnostic Interview Schedule for Children – Revised Version (DISC-R): III. Concurrent criterion validity. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 658-665.

Pliszka, S., Bernet, W., Bukstein, O., & Walter, H. (2007). Practice parameter for the assessment and treatment of children and adolescents with Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 894-921.

Power, T.J., Costigan, T.E., Leff, S.S., Eiraldi, R.B., & Landau, S. (2001). Assessing ADHD across settings: Contributions of behavioral assessment to categorical decision making. *Journal of Clinical Child Psychology, 30*, 399-412.

Power, T.J., Doherty, B.J., Panichelli-Mindel, S.M., Karustis, J.L., Eiraldi, R.B., Anastopoulos, A.D., DuPaul, G.J. (1998a). The predictive validity of parent and teacher reports of ADHD symptoms. *Journal of Psychopathology and Behavioral Assessment, 20*, 57-81.

Power, T.J., Andrews, T.J., Eiraldi, R.B., Doherty, B.J., Ikeda, M.J., DuPaul, G.J., & Landau, S. (1998b). Evaluating Attention Deficit Hyperactivity Disorder using multiple informants: The incremental utility of combining teacher with parent reports. *Psychological Assessment, 10*, 250-260.

Reid, R., & Magg, J.W. (1994). How many fidgets in a pretty much: A critique of behavior rating scales for identifying students with ADHD. *Journal of School Psychology, 32*, 339-354.

Roberts, R.E., Solovitz, B.L., Chen, Y., & Casat, C. (1996). Retest stability of DSM-III-R diagnoses among adolescents using the Diagnostic Interview Schedule for Children-IV (DISC-2.1C). *Journal of Abnormal Child Psychology, 24*, 349-362.

Rowland, A.S., Umbach, D.M., Catoe, K.E., Stallone, L., Long, S., Rabiner, D., et al.

(2001). Studying the epidemiology of attention-deficit hyperactivity disorder. *The Canadian Journal of Psychiatry, 46,* 931-940.

Sayal, K., Letch, N., & Abd, S.E. (2008). Evaluation of screening in children referred for an ADHD assessment. *Child & Adolescent Mental Health, 13*, 41-46.

Schaughency, E.A., & Rothlind, J. (1991). Assessment and classification of attention deficit hyperactivity disorders. *School Psychology Review, 20*, 187-103.

Shaffer, D., Fisher, P., Lucas, C., Dulcan, M., & Schwab-Stone, M. (2000). NIMH Diagnostic Interview Schedule for Children-IV version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry 39*, 28-38.

Simonsen, B.M., & Bullis, M.D. (2007). The effectiveness of using a multiple gating procedure to discriminate among ADHD subtypes. *Journal of Emotional and Behavioral Disorders, 15*, 223-236.

Swanson, J.M. (1992). *School-based assessment and interventions for ADD students.* Irvine, CA: KC Publishing.

Tripp, E.A., & Clarke, B. (2006). Parent and teacher rating scales in the evaluation of attention-deficit hyperactivity disorder: Contribution to diagnosis and differential diagnosis in clinically referred children. *Journal of Developmental and Behavioral Pediatrics, 27*, 209-228.

Volpe, R., DiPerna, J.C., Hintze, J.M., & Shapiro, E.S. (2005). Observing students in

classroom settings: A review of seven coding schemes. *School Psychology Review, 34*, 454-474.

Wender, E. (2004). ADHD Symptoms and parent-teacher agreement. *Journal of Developmental and Behavioral Pediatrics, 25*, 48-49.

Wolraich, M.L., Feurer, I.D., Hannah, J.N., Baumgaerel, A., & Pinnock, T.Y. (1998). Obtaining systematic teacher reports of Disruptive Behavior Disorderss utilizing DSM-IV. *Journal of Abnormal Child Psychology, 26*, 141-152.

Wolraich, M.L., Lambert, E.W., Bickman, L., Simmons, T., Doffing, M.A., & Worley, K.A. (2004). Assessing the impact of parent and teacher agreement on diagnosing Attention-Deficit Hyperactivity Disorder. *Journal of Developmental and Behavioral Pediatrics, 25*, 41-47.

Woodcock, R.W., McGrew, K.S., Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.