ORIGINAL PAPER

Estimates of the Validity and Utility of the Conners' Continuous Performance Test in the Assessment of Inattentive and/or Hyperactive-Impulsive Behaviors in Children

**Mark C. Edwards · Eunice S. Gardner ·
John J. Chelonis · Eldon G. Schulz · Rebecca A. Flake ·
Pamela F. Diaz**

Abstract This study evaluated the validity and classification utility of the Conners' Continuous Performance Test (CCPT) in the assessment of inattentive and hyperactive-impulsive behaviors in children. Significant, positive correlations between the CCPT parameters and behavioral ratings of ADHD behaviors were hypothesized. In addition, it was hypothesized that the CCPT parameters would perform better than a random test (chance) and show fair to moderate utility of classification across the different indices. Participants were 104 children between 6 and 12 years of age who were referred for evaluation of attention problems. The first hypothesis was not supported. There were no significant, positive correlations between the CCPT parameters and parent and teacher ratings of inattentive and hyperactive-impulsive behaviors. The second hypothesis was only partially supported. The CCPT Overall Index and the Omission Errors (84th percentile cutoff) performed better than a random test; however, the utility of the CCPT Overall Index only ranged from poor to slight. Receiver operating characteristic analyses showed the accuracy of the CCPT to be low. The implications and limitations of this study and future research directions are discussed.

Keywords CPT · ADHD · Conner's continuous performance task · Assessment · Children

M. C. Edwards · E. S. Gardner · J. J. Chelonis · E. G. Schulz
Department of Pediatrics,
University of Arkansas for Medical Sciences,
Little Rock, AR 72202-3591, USA

E. S. Gardner · J. J. Chelonis · R. A. Flake · P. F. Diaz
Department of Research and Sponsored Programs,
University of Arkansas at Little Rock, USA

J. J. Chelonis
National Center for Toxicological Research,
Jefferson, Arkansas, USA

R. A. Flake
Department of Psychology, University of Kentucky
at Lexington, USA

E. S. Gardner
Department of Psychology,
Idaho State University at Pocatello, USA

M. C. Edwards (✉)
Pediatric Psychology, Arkansas Children's Hospital,
800 Marshall Street, Slot 512-21, Little Rock, AR 72202-3591
e-mail: edwardsmark@uams.edu

Children with symptoms of Attention-Deficit/Hyperactivity Disorder (ADHD) are among the most common childhood mental health referrals. Using current clinical diagnostic criteria (DSM-IV), prevalence rates in the United States ranged from 7.4% to 9.9% (Barkley, 2006). Developing standards of practice recommend that a variety of methods and informants be considered in the assessment of ADHD, including clinical interviews (parent and child), parent and teacher rating scales, and a physical exam (American Academy of Child and Adolescent Psychiatry, 1991; American Board of Pediatrics, 2000). The subjective nature of many of these methods makes them vulnerable to clinician and informant biases. As a result, there has been interest in identifying or creating objective laboratory measures, such as continuous performance tests, structured behavioral observation systems, the stop signal task, or the go-no go task, that could enhance current assessment measures (Nichols & Waschbusch, 2004; Rapport, Chung, Shore, Denney, & Isaacs, 2000).

The continuous performance test (CPT) is one of the most popular objective measures of ADHD behaviors used in research and marketed for clinical use (Ballard, 1996; Corkum & Siegel, 1993; Epstein, Conners, Sitarenios, &

Erhardt, 1998; McGee, Clark, & Symons, 2000; Nichols & Waschbusch, 2004). The CPT is a vigilance task that requires the participant to remain attentive to changing stimuli typically presented on a computer screen and to respond by pressing a key when specific stimuli appear. CPTs are thought to be cognitively demanding and sensitive to the detection of symptoms of ADHD (Ballard, 1996); however, whether or not the CPT has utility as an ADHD diagnostic instrument has yet to be resolved (Barkley, 1991; Halperin, Sharma, Greenblatt, & Schwartz, 1991; Halperin, Wolf, Greenblatt, & Young, 1991; Losier, McGrath, & Klein, 1996; McGee et al., 2000; Trommer, Hoeppner, Lorber, & Armstrong, 1988).

A contributing factor in the uncertainty of the CPT as a diagnostic instrument is the uncertainty about what processes the CPT measures and how these processes relate to ADHD (Barkley, 1991; Corkum & Siegel, 1993; Halperin, Sharma et al., 1991; Halperin, Wolf et al., 1991; McGee et al., 2000; Rapport et al., 2000). Some research has shown that children with ADHD have impaired performance on the CPT compared to children without ADHD (Epstein, Erkanli, Conners, Klaric, Costello, & Angold, 2003; Halperin, Matter, Bedi, Sharma, & Newcorn, 1992; Losier, et al., 1996; O'Dougherty, Nuechterlein, & Drew, 1984). Other research has shown either weak (DuPaul, Anastopoulus, Shelton, Guevremont, & Metevia, 1992) or no performance differences (Aylward, Verhulst, & Bell, 1990; Corkum & Siegel, 1993, McGee et al., 2000). The CPT has failed to reliably discriminate between children with ADHD and children who have other psychiatric or physical disorders (Halperin et al., 1992; Klee & Garfinkel, 1983; Koelega, 1995; Rovet & Hepworth, 2001; Shapiro & Garfinkel, 1986; Zahn, Kruesi, & Rapoport, 1991). One study (O'Brien, Halperin, Newcorn, Sharma, Wolf, & Morganstein, 1992) was able to find differences in CPT performance between children with Conduct Disorder (CD) and children with ADHD or children with combined ADHD and CD.

The changing ADHD criteria between DSM-III-R and DSM-IV, as well as the differing means of assessing ADHD status makes it difficult to compare results across studies. For example, in studies that used DSM-IV ADHD criteria, one study classified subjects clinically, with no specification of any methods or algorithm used (Schatz, Ballantyne, & Trauner, 2001). Another study classified subjects based on a structured diagnostic interview of parents alone (Epstein et al., 2003). A third study classified subjects based on diagnosis by a clinical multidisciplinary team, scores above 1.5 standard deviations on parent and teacher rating scales measuring ADHD behaviors, and blind review by a clinical psychologist (McGee et al., 2000). None of these studies subtyped ADHD for their primary analyses. These differing criteria for identifying ADHD can result in dissimilar samples, making comparisons across study results difficult.

Another difficulty in comparing results across studies is the differing versions of the CPT. The differing parameters of the CPT include the type of stimuli presented (e.g., numbers, letters, symbols, auditory), proportion of target (requires a response) to non-target stimuli presented, duration the stimulus is presented, interval between stimuli, and duration of the task. Three versions of the CPT that have been marketed for clinical use include the Gordon Diagnostic System Vigilance Task (GDS; Gordon, 1986), the Test of Variables of Attention (TOVA; Dupuy & Greenberg, 1993), and the Conners' Continuous Performance Test (CCPT; Conners, 1994). The GDS follows the "rare target" (Nichols & Waschbusch, 2004) paradigm set forth by Rosvold, Mirsky, Sarason, Bransome, and Beck (1956) and contains 10% targets and 90% non-targets. In this paradigm, the participants must attend to the changing stimuli and respond infrequently when the target stimuli appear. The CCPT uses a "response inhibition" paradigm which includes 90% target stimuli and 10% non-target stimuli while changing the interstimulus interval in a counter-balanced manner across blocks of time. In this paradigm, the subject is frequently responding and must inhibit a response when presented with the non-target stimuli. The TOVA utilizes both paradigms by presenting 22.5% targets and 77.5% non-targets in the first half of the test and then switching to 77.5% targets and 22.5% non-targets in the last half of the test (Dupuy & Greenberg, 1993).

There are some conceptual differences between these two paradigms. The rare target paradigm is thought to assess sustained attention, whereas, the response inhibition paradigm is thought to assess behavioral inhibition. Tests based on this later paradigm, like the CCPT, are thought to hold promise in the assessment of ADHD as they are consistent with a current theory which sees a delay in behavioral inhibition as being the primary deficit in ADHD (Barkley, 1997). Such differences in the CPT parameters may be important as the research and clinical literature suggests that the current subtypes of ADHD represent heterogeneous groups and perhaps distinct disorders. For example, within the Predominantly Inattentive Type, Barkley (2006) suggests that a subset of these children (30% to 50%) have a cluster of symptoms (e.g., hypoactivity, slow cognitive processing speed, passivity, daydreaming) called sluggish cognitive tempo, while another subset may have a mild version of ADHD Combined Type with subthreshold hyperactive-impulsive symptoms. Barkley suggests that additional research may reveal these subsets to be two distinct disorders. It is possible that different objective measures, as well as different CPT paradigms, will be needed to assess the different aspects of ADHD.

The utility of the GDS, TOVA, and CCPT for diagnosing children with ADHD has yet to be satisfactorily demonstrated. The GDS has not consistently been found to be sensitive to ADHD diagnosis. Some studies have found differences between children with ADHD and children without

ADHD (Aylward et al., 1990) while others have failed to find differences (DuPaul et al., 1992; Trommer et al., 1988). Most of the studies reviewed above assessed differences between groups. Such studies are less informative to clinical practice, which is primarily concerned with issues of diagnostic accuracy for individual patients. The TOVA has had more consistent results in identifying patients with ADHD; however, it has been found to lack specificity, misidentifying as many as 35% of those without ADHD (Forbes, 1998; Schatz et al., 2001). Research using the CCPT found that it correctly identified children and adults with ADHD (sensitivity) only 52% and 55% of the time, respectively (Epstein et al. 1998; McGee et al., 2000). However, several of the CCPT performance measures have been found to have significant relationships to ADHD symptomology (Epstein et al., 2003).

This study seeks to examine the validity and utility of the CCPT in the assessment of children with symptoms of ADHD. This study examines the CCPT using a relatively large sample size ($N = 104$). In addition, previous research and reviews have identified the use of community (non-clinically referred) controls as potentially inflating differences between those with ADHD and those without (Corkum & Siegel, 1993; McGee et al., 2000). This study uses a sample of children clinically referred for ADHD evaluation. Furthermore, this study included both the inattentive and hyperactive-impulsive subtypes in the primary analyses.

The first purpose of this study was to determine the validity of the CCPT relative to other methods of assessing ADHD symptoms in children. Specifically, the associations between the CCPT Overall Index, Omissions, and Commissions and standardized parent and teacher behavioral ratings were examined. It was hypothesized that there would be significant, positive correlations between the CCPT parameters and parent and teacher ratings of inattentive and hyperactive-impulsive behaviors. The second purpose of this study was to examine the utility of the CCPT in correctly classifying children with and without problems with inattention and hyperactive-impulsive behaviors. This was done by looking at several indices used to assess the utility of measures, including predictive power, efficiency, and receiver operating characteristics. It was hypothesized that the CCPT parameters would perform better than a random test (chance) and show fair to moderate utility of classification across the different indices.

Methods

Participants

Participants were 104 children between 6 and 12 years of age (Mean age: 8.77, *SD*: 1.98) who were recruited from consecutive referrals to an outpatient developmental center at Arkansas Children's Hospital (pediatrics department). A breakdown of payment source for patients seen at this center includes Medicaid (57.6%), commercial insurance (42%) and uninsured (0.4%). Participants were 78 males (75 %) and 26 females (25 %). Race of participants included 85 White (81.7 %), 16 Black (15.4 %), 2 Hispanic (1.9 %), and 1 American Indian (1.0 %). Ninety five (91.3%) of the participants had no history of ADHD medications, four had a previous history of ADHD medication use, and five were currently taking medications. None of the participants were on medications for ADHD on the day of the evaluation. Teacher ratings were not available for four of the participants. The major cognitive and behavioral characteristics of the sample are listed in Table 1. There were no differences across diagnostic categories, using a structured diagnostic interview, on the cognitive measures. There were significant differences across the diagnostic categories on parent and teacher ratings of externalizing and attention problems behaviors, and parent ratings of internalizing behaviors, with the Hyperactive-Impulsive/Combined Type of ADHD having consistently higher scores. As an indicator of possible comorbidities in the sample, Table 2 shows the frequency of participants who met the diagnostic symptom count criteria for the various disorders based on a parent diagnostic checklist. Sixty-five percent of the total sample screened positive for one or more other psychiatric disorders based on a parent checklist, with 40%, 51.6% and 83.7% of the No ADHD, Inattentive Type, and Hyperactive-Impulsive/Combined Type screening positive for at least one additional psychiatric disorder, respectively.

Measures

Cognitive measures

Cognitive measures were used to describe the sample and as a covariate in the correlational analyses. Each participant was administered the Kaufman Brief Intelligence Test (KBIT; Kaufman & Kaufman, 1990). The KBIT is a brief, individually administered measure of verbal and nonverbal intelligence. The split-half reliabilities are .92 and .85, respectively. Correlation of the KBIT IQ Composite with the Full Scale IQ from the Wechsler Intelligence Scale for Children-Revised is .80. Academic achievement was assessed using the Wide Range Achievement Test-3 (WRAT-3; Wilkinson, 1993). Subscales include reading recognition, spelling, and arithmetic. The reliabilities for the three subtests range from .92 to .95. The developer provides item statistics showing that items range in difficulty to support content validity and hypotheses and evidence to support construct validity. An independent study found a positive, significant relationship between the WISC-III and WRAT-3 scores in a sample

Table 1 Mean (SD) subject characteristics by ADHD classification based on the Diagnostic Interview Schedule for Children-Parent Version

Characteristic	Total sample (N = 104)	No ADHD (n = 21)	Inattentive type ADHD (n = 33)	Hyperactive-impulsive & combined type ADHD (n = 50)	F
K-BIT IQ composite[a]	97.07 (13.36)	97.71 (14.66)	97.91 (15.34)	96.24 (11.51)	0.18
WRAT-III reading achievement[a]	93.47 (15.88)	93.57 (17.75)	94.09 (13.76)	93.02 (16.66)	0.04
WRAT-III spelling achievement[a]	92.64 (13.65)	92.76 (14.81)	94.48 (13.41)	91.38 (13.45)	0.51
WRAT-III math achievement[a]	87.62 (14.96)	88.33 (14.11)	88.55 (13.26)	86.72 (16.51)	0.17
CBCL internalizing[b]	60.06 (9.94)	57.05 (9.26)	58.10 (8.26)	62.54 (10.69)	3.26*
CBCL externalizing[b]	62.24 (11.21)	55.57 (9.90)	56.81 (9.11)	68.40 (9.59)	20.68**
CBCL attention problems[b]	69.77 (9.72)	64.62 (11.21)	68.32 (8.17)	72.84 (8.95)	6.40*
TRF internalizing[b]	54.56 (8.98)	52.95 (10.00)	52.48 (9.27)	56.80 (7.91)	2.69
TRF externalizing[b]	57.67 (10.19)	53.50 (8.22)	52.00 (8.72)	63.69 (8.67)	20.52**
TRF attention problems[b]	65.38 (9.14)	60.40 (6.28)	63.36 (9.03)	69.07 (8.92)	8.58**

[a] Standard Scores.

[b] T-scores.

*$p < .05$.

**$p < .01$.

Measurement of child behavior

The ADHD portion of the Computerized-Diagnostic Interview Schedule for Children-4.0 (C-DISC; Columbia University DISC Development Group, 2000) was used as a method of classification based on DSM-IV (American Psychiatric Association, 1994) ADHD criteria. The parent version of the C-DISC is a structured diagnostic interview for children between 6 and 17 years of age. This program follows an algorithm in which the informant is systematically asked a series of closed-ended questions related to specific problem areas, such as the presence and duration of specific psychiatric symptoms. Test-retest reliability for the parent version

of school-age special education students (Vance & Fuller, 1995).

ADHD scale is .79. Criterion validity of the DISC version 2.3 has been reported, with kappa's for ADHD at .72 for the parent version.

The Conners' ADHD/DSM-IV Scales (CADS; Conners, 1997) were used to assess the level of children's ADHD behaviors. The CADS is a 28 item self-report measure that is keyed to DSM-IV ADHD symptoms. The CADS reference group consists of a national sample of 2426 and 1897 (parent and teacher, respectively) children from 3–17 years of age. The Inattentive and Hyperactive-Impulsive subscales were used in this study. Internal consistency of the subscales was reported to range from .82 to .99. No validity data has been reported on the CADS; however, the Inattentive and Hyperactive-Impulsive subscales of the CADS are composed of the DSM-IV symptoms for ADHD. Several similar scales have been developed with some evidence for

Table 2 Frequency (%) of participants that met screening cutoff for specific disorders on the Child Symptom Inventory-4: Parent Checklist

Categories	Total sample (N = 104)	No ADHD (n = 21)	Inattentive type ADHD (n = 33)	Hyperactive-impulsive & combined type ADHD (n = 50)
Specific phobia	41 (39.4%)	7 (33.3%)	10 (30.3%)	24 (48%)
Oppositional defiant disorder	37 (35.6%)	4 (19.0%)	3 (9.1%)	30 (60%)
Conduct disorder	27 (26.0%)	2 (9.5%)	0 (0.0%)	25 (50%)
Obsessions	25 (24.0%)	4 (19.0%)	7 (21.2%)	14 (28%)
Compulsions	12 (11.5%)	4 (19.0%)	3 (9.1%)	5 (10%)
Generalized anxiety disorder	10 (9.6%)	3 (14.3%)	4 (12.1%)	3 (3%)
Separation anxiety	8 (7.7%)	1 (4.8%)	2 (6.1%)	5 (10%)
Asperger's disorder	5 (4.8%)	0 (0.0%)	0 (0.0%)	5 (10%)
Dysthymia	3 (2.9%)	0 (0.0%)	0 (0.0%)	3 (6%)
Autism	2 (1.9%)	0 (0.0%)	0 (0.0%)	2 (4%)
Social phobia	2 (1.9%)	2 (9.5%)	0 (0.0%)	0 (0%)
Schizophrenia	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0%)
Major depressive disorder	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0%)

predictive, convergent, and discriminant validity reported (Pelham, Fabiano, & Massetti, 2005).

In order to further describe the sample, the Child Behavior Checklist (CBCL; Achenbach, 1991a) and Teacher Report Form (TRF; Achenbach, 1991b) were administered to the parents and teachers. Internal consistency of the empirically-based subscales was reported to range from .72 to .97, and test-retest reliabilities ranged from .60 to .95. The manuals contain extensive validity data, including data supporting strong concurrent and discriminant validity (Furlong, 1998).

As a screening for possible psychiatric comorbidities in the sample, the parents were administered the Child Symptom Inventory-4: Parent Checklist (Gadow & Sprafkin, 1997). This checklist assesses the presence and frequency of specific DSM-IV psychiatric symptoms. This checklist yields a screening cutoff score for each diagnostic category.

Conners' continuous performance test (CCPT)

The CCPT is a computerized continuous performance task. The standard version of the CCPT consists of 360 trials in which a single letter appears on a computer screen for approximately 250 ms. The subject is instructed to press the space bar for each letter except the letter X. The percentage of trials in which a letter other than X appears is 90%. The task is approximately 14 minutes in duration. Conners (1994) reported a mean false positive rate of 10.8% and a mean false negative rate of 11.8% using the original reference sample. A cross-validation for 6–17 year olds reported a false positive rate of 13.5% and a false negative rate of 26.1%.

The association between parent and teacher ratings of inattentive behaviors was .245 ($p = .014$) and the association between parent and teacher ratings of hyperactive-impulsive behaviors was .490 ($p < .001$). Table 3 shows the categorical agreement (efficiency; True Positive Rate & True Negative Rate) between the different measures. In general, there was higher agreement between hyperactive-impulsive measures than inattentive measures, higher agreement between parent ratings and the parent diagnostic interview (C-DISC) than teacher ratings and the C-DISC, and higher agreement be-

Table 3 Agreement (efficiency; true positive rate & true negative rate) between measures

			T-score cutoff	
	N	63	65	67
Inattentive				
CADS parent and teacher	100	.64	.65	.60
C-DISC & CADS parent	54	.61	.61	.69
C-DISC & CADS yeacher	53	.57	.53	.53
Hyperactive-impulsive				
CADS parent and teacher	100	.60	.63	.67
C-DISC & CADS parent	71	.87	.90	.90
C-DISC & CADS teacher	67	.62	.61	.64

tween most measures with higher T-score cutoffs than lower cutoffs.

Procedures

All patients who were within the 6 to 12 year age range and were being evaluated for a suspected attention problem were approached for recruitment into the study. Though the majority of patients consented to participate, a minority of patients did not. No data was collected on the patients that were not consented, consequently, differences between participants and non-participants could not be assessed. Participants were consecutively recruited and informed consent was obtained during their clinical evaluation at a developmental center. As a routine part of the clinic appointment, all participants were mailed behavioral rating scales to be completed by a parent and teacher and brought with them to their appointment. The parent forms were completed by a caregiver, in most cases the mother, but in some cases the father or guardian. All other measures used in this study were not routinely part of the clinic appointment and were administered specifically for this study by trained research assistants (primarily undergraduate students) and supervised by a licensed clinical psychologist. The research assistants used standardized administration procedures as detailed in the respective manuals. The parent was administered the ADHD portion of the C-DISC while the child was administered the cognitive measures and the CCPT. All testing was done in private clinic rooms.

Statistical analyses

Pearson correlations were computed to assess the relationship between the CCPT Overall Index, Omissions, and Commissions scores and the Inattentive and Hyperactive-Impulsive subscales from the parent and teacher CADS. The CCPT Overall Index and Omissions scores were significantly and negatively correlated with IQ (rs (104) $= -.238$ & $-.279$, $p < .015$, respectively). The Omissions score was significantly and negatively correlated with age (r (104) $= -.346$, $p < .001$). Therefore, age and IQ were partialed out where appropriate. The guidelines for assessing the magnitude of association proposed by Cohen (1988) were used: $.1 =$ small, $.3 =$ medium, and $.5 =$ large.

Several parameters were used to describe the utility of the CCPT, including positive (PPP) and negative predictive power (NPP), quality of efficiency (κ), and receiver operator characteristics. In this study, PPP refers to the proportion of participants who test positive on a CCPT parameter (Overall Index, Omissions, Commissions) who have a problem with inattentive or hyperactive-impulsive behaviors based on the classification approaches described below. Conversely, NPP refers to the proportion of participants who test negative on a

CCPT parameter who do not have a problem with inattentive or hyperactive-impulsive behaviors. Sensitivity and specificity are parameters which are frequently used to judge a measure's ability to predict classification within a population with or without a specific condition, respectively. However, PPP and NPP are more intuitive parameters for diagnostic situations than are sensitivity and specificity because in a diagnostic situation the presence or absence of a condition is not known. However, interpretation of PPP and NPP is not straightforward. Assessment of the magnitude of the predictive values depends on their position relative to their range of values. The effective range of PPP is from P (prevalence of a positive identification in the sample) to 1, while the range for NPP is from P' to 1. In order to facilitate interpretation of the predictive values, PPP and NPP were recalibrated, taking into account their range, to yield corrected scores that range from 0 (random test) to 1 (perfect test; cPPP and cNPP, respectively; see Kraemer, 1992 for formulas). In addition, confidence intervals (95%) were calculated to assess the accuracy of the estimated predictive values given the sample sizes. If the lower bound of the interval falls below zero, this indicates that the observed predictive value is not significantly above zero at the 5% significance level.

Two parameters were used to evaluate the accuracy of the CCPT in classifying problems with inattentive or hyperactive-impulsive behaviors. A kappa coefficient (κ) was used as an index of the quality of the CCPT's efficiency. Efficiency is defined as the probability that classification based on the CCPT and classification based on other measures of inattentive or hyperactive-impulsive behaviors agree (True Positive Rate + True Negative Rate; Kraemer, 1992). Like PPP and NPP, efficiency is an uncalibrated measure that is dependent on both the prevalence (P) and level of the test (Q). Kappa recalibrates efficiency to facilitate interpretation, yielding a parameter ranging from 0, indicating no agreement, to 1, indicating total agreement. In addition, given the concern about the accuracy of the kappa estimate given the sample sizes, a statistical jackknife estimation procedure (Kraemer, 1992) was used to obtain a confidence interval that indicates the margin of error. If the lower bound of the interval falls below zero, this indicates that the observed kappa is not significantly above zero at the 5% significance level. The following benchmarks proposed by Landis and Koch (1977) were used to interpret cPPP, cNPP, and κ: 0 = poor; 0 to .2 = slight; .21 to .40 = fair; .41 to .60 = moderate; .61 to .80 = substantial and .81 to 1 = near perfect.

The receiver operating characteristics curve (ROC) was also used to assess the classification accuracy of the CCPT. ROC analysis yields an index of area under the curve (AUC). In this case, the AUC statistic represents the probability that a randomly chosen subject that has been classified as having a problem with inattentive or hyperactive-impulsive behaviors will score higher on the CCPT parameter than a randomly

chosen subject that does not have a problem. The AUC parameter varies from .50 (no better than chance prediction) to 1.00 (perfect rate of prediction). An advantage of ROC analysis and the AUC parameter is that it is not dependent on prevalence or on the cutoff score.

Classification criteria

Many of the analyses described above require that each case be categorized into a 2×2 contingency table. Several factors present a challenge for such categorization, including the continuous variables of the CCPT, dimensional nature of ADHD symptoms, overlapping symptoms in the current subtyping of ADHD (i.e., inattention), and use of multiple methods and informants. These factors require the establishment of cutoffs and algorithms for integrating results from different methods and settings.

A score of 9 or above on the CCPT Overall Index was used to positively classify participants as having a problem with inattentive or hyperactive-impulsive behaviors, as suggested by the manual. For the Omissions and Commissions indices, cutoffs of the 84th and 90th percentiles were used, similar to the CCPT interpretative guidelines. For the analysis of the CADS data, T scores were used to adjust for age and gender variations. Establishment of cutoffs is somewhat arbitrary and where the cutoff is set will affect identification rates. Epidemiological studies have used cutoff scores ranging from 1 to 2 standard deviations to represent a child's behavior as "developmentally inappropriate." Barkley (2006) discussed this issue, citing data and concerns that a 1.5 standard deviation above the mean criterion may over identify school-aged children as hyperactive and a 2 standard deviation above the mean cutoff may under identify. He advocates for a cutoff "somewhat below" 2 standard deviations. In this study, cutoffs of 1.3, 1.5, and 1.7 were used.

This study used two approaches for identifying problems with inattentive and hyperactive-impulsive behaviors. The first approach employed a structured diagnostic interview (C-DISC) with the parent. The C-DISC is keyed to DSM-IV criteria, the diagnostic system used by most clinicians and researchers. DSM-IV criteria are considered the "most rigorous and most empirically derived criteria ever available" (Barkley, 2006, p. 84). DSM-IV based structured diagnostic interviews tend to have higher reliability than less structured methods, though they may not add much incremental validity over parent and teacher rating scales (Pelham et al., 2005). Using this approach, participants were classified as having a problem with inattentive or hyperactive-impulsive behaviors if they had a positive "diagnosis" on the C-DISC. Subgroups included No ADHD ($n = 21$), Inattentive Type ($n = 33$), Hyperactive-Impulsive Type ($n = 2$) and Combined Type ($n = 48$). Given the low number in the Hyperactive-Impulsive Type, they were not analyzed

separately. The use of a structured diagnostic interview by itself to identify ADHD is limited by its reliance on a single method and informant and is strictly a categorical approach, as opposed to normative or dimensional. However, a parent interview is a reasonable comparison for the CCPT as it is a method that is recommended by all existing standards of practice (American Academy of Child and Adolescent Psychiatry, 1991; American Board of Pediatrics, 2000; Barkley, 2006).

A second approach to classification utilized both categorical and dimensional measures (Jensen & Watanabe, 1999) and multiple informants. In this multiple criteria approach, participants were classified as having a problem with inattentive or hyperactive-impulsive behaviors based on a positive classification of ADHD on the C-DISC and scores at or above a cutoff on the relevant subscales of the parent and teacher rating scales (CADS; e.g., above cutoff on the CADS Hyperactive-Impulsive subscales for the C-DISC Hyperactive-Impulsive Type). Criteria for the Combined subgroup included a positive C-DISC Combined Type and scores above the cutoffs on either the Inattentive or Hyperactive-Impulsive subscales on both parent and teacher CADS. Subgroups included Inattentive ($n = 12$–15, depending on cutoff), Hyperactive-Impulsive ($n = 1$), Combined ($n = 32$–36), and None ($n = 21$). All other cases ($n = 27$–34) were considered subthreshold cases (e.g., participants with a positive classification of ADHD on the C-DISC and parent or teacher ratings below the cutoff).

In order to replicate a clinical situation, all cases were included in all the analyses. The CCPT was evaluated against classification groups based on symptoms types. Groups included All Subtypes (inattentive, hyperactive-impulsive, combined), Inattentive/Combined, Hyperactive-Impulsive/Combined, and Inattentive Only. Positive cases for each group were based on the criteria described above. Negative cases included those that did not meet criteria for a subgroup, those that met criteria for a subgroup different than was being examined (e.g., hyperactive-impulsive when inattentive subgroup was being analyzed), and subthreshold cases. Since inattentive symptoms are shared across the Inattentive and Combined subtypes, analyses are presented using the Inattentive subgroup only and combining the Inattentive and Combined subgroups.

Results

Concordance between the CCPT and parent and teacher ratings

There were no significant correlations between the CCPT Overall Index score and parent and teacher ratings of inattentive and hyperactive-impulsive behaviors with IQ partialled

out (partial rs ranged from .043 to .096). The magnitude of the associations can be considered small. Statistical power ranged from .10 to .24. With the study sample size, a correlation of .248 (small to medium size) would be needed to detect a difference between the null and alternative hypotheses at the .05 level with .80 power. There were no significant correlations between the CCPT Omissions score and parent and teacher ratings of inattentive and hyperactive-impulsive behaviors with IQ and age partialled out (partial rs ranged from .090 to .194). There were no significant correlations between the CCPT Commissions score and parent ratings of inattentive and hyperactive-impulsive behaviors (rs $= -.100$ and $-.114$, respectively). There were significant negative correlations (unexpected) between the CCPT Commissions score and teacher ratings of inattentive and hyperactive-impulsive behaviors (rs $(100) = -.204$ & $-.274$, ps $= .042$ and .006, respectively).

Accuracy and classification utility

Table 4 shows the parameters used to describe the classification utility and accuracy of the CCPT Overall Index. Most participants (58.7% to 61.0%; 1-Q) performed well or "passed" the CCPT (using the recommended Overall Index cutoff of 9). The notable differences between the predictive values (PPP, NPP) and their recalibrated values (cPPP, cNPP) is a function of the differing prevalence rates (P), which varied from 12.0% to 79.8% depending on the criteria used to classify participants. This adjustment allows for a more straightforward interpretation by yielding a range from 0 (random test) to 1 (perfect test).

The quality of efficiency of the CCPT Overall Index ranged from poor to slight. The CCPT Overall Index was not reliably better than a random test in correctly classifying children in the Inattentive Only classification group. The CCPT Overall Index showed slight quality of efficiency in the other classification groups, ranging from .088 to .199. The positive predictive power of the CCPT Overall Index ranged from poor to fair. The cPPP was fair for the All Subtypes DISC Only, All Subtypes Multiple Criteria (67 cutoff), and Inattentive/Combined DISC Only classifications. It showed no reliable positive predictive power in identifying children with inattention only. The CCPT Overall Index showed slight positive predictive power in the other groups, ranging from .116 to .194. The cNPP of the CCPT Overall Index was fair for the Hyperactive-Impulsive/Combined classification (65 CADS cutoff). The CCPT Overall Index showed no reliable negative predictive power in identifying children in the Inattention Only group. The negative predictive power was slight for all the other classification groups, ranging from .053 to .205.

Although the results were similar across classification groups, in general, the quality of efficiency (κ) and

Table 4 Parameters to describe the accuracy and classification utility of the CCPT overall index (9.0 cutoff)

Classification	N	P^a	Q^b	PPP^c	NPP^d	$cPPP^e$ (CI^f)	$cNPP^g$ (CI)	κ^h (CI)	AUC^i
All subtypes									
DISC only	104	.798	.413	.860	.246	.309 (±.089)	.055 (±.044)	.094 (.072 − .118)	.644*
Multiple criteria 67 CADS cutoff^j	100	.440	.390	.564	.639	.222 (±.081)	.180 (±.075)	.199 (.177 − .225)	.605
Multiple criteria 65 CADS cutoff	100	.470	.390	.564	.590	.178 (±.075)	.128 (±.065)	.149 (.126 − .174)	.580
Multiple criteria 63 CADS cutoff	100	.500	.390	.564	.541	.128 (±.066)	.082 (±.059)	.100 (.078 − .125)	.563
Inattentive/combined									
DISC only	104	.779	.413	.837	.262	.264 (±.085)	.053 (±.043)	.088 (.067 − .112)	.637*
Multiple criteria 67 CADS cutoff^k	100	.370	.390	.487	.705	.186 (±.076)	.202 (±.079)	.194 (.170 − .221)	.601
Multiple criteria 65 CADS cutoff	100	.400	.390	.487	.656	.145 (±.069)	.139 (±.068)	.142 (.118 − .168)	.574
Multiple criteria 63 CADS cutoff	100	.420	.390	.487	.623	.116 (±.063)	.102 (±.059)	.109 (.085 − .134)	.554
Hyperactive-impulsive/combined									
DISC only	104	.481	.413	.581	.590	.194 (±.076)	.148 (±.068)	.168 (.147 − .192)	.592
Multiple criteria 67 CADS cutoff	100	.320	.390	.436	.754	.170 (±.074)	.232 (±.083)	.196 (.171 − .224)	.593
Multiple criteria 65 CADS cutoff	100	.330	.390	.436	.738	.158 (±.071)	.205 (±.079)	.179 (.153 − .206)	.576
Multiple criteria 63 CADS cutoff^l	100	.350	.390	.436	.705	.132 (±.066)	.157 (±.071)	.144 (.118 − .170)	.571
Inattentive only									
DISC only	104	.317	.413	.279	.656	−.056	−.085	−.068	.501
Multiple criteria 67 CADS cutoff^m	100	.120	.390	.128	.885	.009 (±.019)	.044 (±.040)	.015 (−.022 − .049)	.552
Multiple criteria 65 CADS cutoff	100	.140	.390	.128	.852	−.014	−.054	−.022 (−.059 − .011)	.525
Multiple criteria 63 CADS cutoff	100	.150	.390	.128	.836	−.026	−.093	−.040	.497

*Significantly better than chance ($p < .05$).

$^a P$ = Prevalence of a positive identification of problems.

$^b Q$ = Level of the test (Base rate of participants with a CCPT Overall Index score at or above the cutoff).

^{c}PPP = Positive Predictive Power: proportion of participants who test positive for ADHD symptoms on the CCPT Overall Index who were classified as having a problem.

^{d}NPP = Negative Predictive Power: proportion of participants who test negative for ADHD symptoms on the CCPT Overall Index who were classified as having no problem.

^{e}cPPP = PPP recalibrated to correct for prevalence, yields score that ranges from 0 (random test) to 1 (perfect test).

^{f}95% confidence interval.

^{g}cNPP: NPP recalibrated to correct for prevalence; yields score that ranges from 0 (random test) to 1 (perfect test).

$^h \kappa$ = Kappa statistic; Quality (recalibration) of efficiency. Efficiency is defined as the probability that classification based on the CCPT and classification based on other measures of ADHD behaviors agree (True Positive Rate + True Negative Rate; 0 = no agreement other than that which would be expected by chance; 1.00 = total agreement).

^{i}AUC = Area under the ROC curve: Index of identification accuracy (.50 = no better than chance prediction; 1.00 = perfect rate of prediction).

^{j}Identified group = positive DISC any subtype & positive parent and teacher CADS; Non-identified group = negative DISC or subthreshold cases.

^{k}Subthreshold and hyperactive-impulsive cases included in non-identified group.

^{l}Subthreshold and inattentive only cases included in non-identified group.

^{m}Subthreshold and hyperactive-impulsive/combined cases included in non-identified group.

quality of the negative predictive power (cNPP) was higher for the multiple criteria classifications than the C-DISC alone. In addition, the quality of efficiency was higher for the higher CADS cutoff scores in the multiple criteria groups. There was higher quality of the positive predictive power (cPPP) when the C-DISC alone was used for classification than the multiple criteria.

Based on the ROC analyses, the CCPT Overall Index was able to correctly classify participants with and without problems with inattentive and hyperactive-impulsive behaviors significantly better than chance for only two of the 16 clas-

sifications. The CCPT predicted identification status for the All Subtype DISC Only and for the Inattentive/Combined DISC Only classifications significantly better than chance (AUC = .644, .637, respectively). The power in the various non-significant samples ranged from 0% to 68%. The smallest effect size that could be detected at 80% power given the sample sizes in was .623.

Analyses using the Omission Errors as the CCPT parameter for identification yielded slight to fair quality of efficiency. The Omission score with the 84th percentile cutoff showed improved quality of efficiency over the Overall Index

when the C-DISC was used for classification. The Commission scores and the Omission (90th percentile cutoff) scores showed poor to slight quality of efficiency across all the classification approaches (see Table 5).

Discussion

The first hypothesis that there would be a significant, positive correlation between the CCPT parameters and parent and teacher ratings of inattentive and hyperactive-impulsive behaviors was not supported. A previous study using the CCPT with children also found no significant associations between the Overall Index and parent and teacher ratings (Mcgee et al., 2000). Previous studies examining the association between CPT indices and behavior ratings have been inconsistent. Some have found positive correlations (e.g., Klee & Garfinkel, 1983), while others have found no significant correlations (e.g., DuPaul et al., 1992). Reasons for such inconsistent findings could be related to sampling and methodological variations, as well as a possible poor correspondence between rated behaviors and constructs measured by the CPT. It is not unusual for laboratory measures to have weak or no correlations with behavioral measures of inattention and hyperactivity-impulsivity (Barkley, 1991). In addition, McGee and colleagues (2000) suggest that such negative findings in clinical samples may be the result of the restriction of range inherent in parent and teacher ratings; most clinic-referred children, regardless of diagnoses, have high ratings on such measures.

The second hypothesis that the CCPT parameters would perform better than a random test (chance) and show fair to moderate utility of classification across the different CCPT indices was only partially supported. For the most part, the CCPT Overall Index and the Omission Errors (84th percentile cutoff) did perform better than a random test. However, the utility of the CCPT Overall Index as measured by the quality of efficiency (κ) only ranged from poor to slight, depending on the classification group. The accuracy in identifying cases of inattentive and hyperactive-impulsive problems (cPPP) was fair for three of the 16 classification groups. The accuracy in identifying cases without inattentive and hyperactive-impulsive problems (cNPP) was fair for one of the 16 classification groups. No clear pattern emerged based on classification groups, with the exception of consistently poor accuracy (κ) in correctly identifying problems with and without inattention only. The accuracy of the CCPT as assessed by the AUC statistic can be considered low, as indicated by a range from .50 to .64. Overall, the results of this study suggest that the CCPT has limited utility in the assessment of inattentive and hyperactive-impulsive behaviors in clinic referred children.

There are several possible explanations for the generally low utility of the CCPT found in the present study. This study primarily used the Overall Index for analysis of the diagnostic utility of the CCPT. The developer advocates the use of this parameter and it is likely the one used by clinicians to interpret the CCPT results for diagnostic purposes. The CCPT manual recommends that a score of 9 or above on the index can be used to suggest an attentional problem. The CCPT narrative report uses a trichotomous classification system where scores of less than 8 are indicative of no problem, scores between 8 and 11 are considered unclear cases, and scores greater than 11 offered the strongest evidence of an attentional problem. Additional research can explore the issue of optimal cutoff for the CCPT Overall Index and the other CCPT scores.

Another possible reason for the findings differing from the a priori predictions may be the issue of what the CCPT is measuring. As reviewed in the introduction, the CPT appears to be a multidimensional measure that is sensitive to a variety of disorders. For example, in a recent study using the CCPT (McGee et al., 2000), the Overall Index was associated with a measure of phonological skills and was unable to discriminate between groups of children with ADHD and reading disorders. The authors speculated that the CCPT's use of both letters as stimuli and continuous responding confounds phonological skills with inhibition.

The heterogeneous nature of clinic referred samples may affect the precision of the CCPT in identifying ADHD. Psychiatric comorbidities tend to be higher in clinic referred samples than community based samples (Szatmari, Offord, & Boyle, 1989; Pfiffner et al., 1999; Wilens et al., 2002), with as many as 80% of samples of school-aged children with ADHD having at least one other disorder. The rates of possible comorbidities based on a parent checklist in this sample found a similar result (65% for the total sample; 83.7% for the Hyperactive-Impulsive/Combined subtype). The small sample size and lack of thorough measurement of comorbidities in the present study precluded further exploration of the construct validity of the CCPT and examination of the diagnostic performance of the CCPT in subgroups of participants with ADHD. In addition, the small sample size precluded subgroup analyses based on gender, an important and current area of research in ADHD. Future studies will need to address the issues of what constructs the CCPT may be measuring, and how different subgroups of children with ADHD perform on the CCPT.

The findings do provide some limited support for the notion that measures like the CCPT that use the response inhibition paradigm may be best when behavioral inhibition is an issue. Classification groups which included participants with hyperactive-impulsive problems were more accurate than groups with inattention only. Other tests may be more appropriate for use with the inattentive subtypes of ADHD.

Table 5 Parameters to describe the accuracy and identification utility of the CCPT omission and commission errors

Measures	N	P[a]	Q[b]	PPP[c]	NPP[d]	cPPP[e] (CI[f])	cNPP[g] (CI)	κ[h] (95% CI)	AUC[i]
Percentile score of 84 as cutoff for CCPT									
Omission errors									
DISC only all subtypes	104	.798	.625	.846	.282	.238 (±.082)	.100 (±.058)	.141 (.111 − .171)	.593
Multiple criteria all subtypes[j]	100	.440	.610	.492	.641	.093 (±.057)	.184 (±.076)	.123 (.102 − .148)	.564
DISC only inattentive/combined	104	.779	.625	.831	.308	.235 (±.081)	.111 (±.069)	.151 (.112 − .180)	.594
Multiple criteria inattentive/combined[k]	100	.370	.610	.410	.692	.063 (±.048)	.168 (±.073)	.092 (.071 − .116)	.552
DISC only hyperactive-impulsive/ combined	104	.481	.625	.569	.667	.170 (±.072)	.307 (±.089)	.219 (.199 − .244)	.611
Multiple criteria hyperactive-impulsive/combined[l]	100	.320	.610	.393	.795	.108 (±.061)	.359 (±.094)	.166 (.146 − .191)	.603
Commission errors									
DISC only all subtypes	104	.798	.260	.741	.182	−.284	−.025	−.046	.454
Multiple criteria all subtypes	100	.440	.260	.423	.554	−.030	−.014	−.019	.491
DISC only inattentive/combined	104	.779	.260	.741	.208	−.172	−.017	−.031	.471
Multiple criteria inattentive/combined	100	.370	.260	.308	.608	−.099	−.059	−.074	.465
DISC only hyperactive-impulsive/ combined	104	.481	.260	.481	.519	.001 (±.007)	.001 (±.004)	.001 (−.099 − .014)	.500
Multiple criteria hyperactive-impulsive/ combined	100	.320	.260	.308	.676	−.018	−.014	−.016	.493
Percentile score of 90 as cutoff for CCPT									
Omission errors									
DISC only all subtypes	104	.798	.462	.833	.232	.175 (±.073)	.038 (±.037)	.062 (.039 − .088)	.550
Multiple criteria all subtypes	100	.440	.450	.422	.545	−.032	−.033	−.032	.484
DISC only inattentive/combined	104	.779	.462	.813	.250	.152 (±.069)	.037 (±.036)	.060 (.037 − .085)	.545
Multiple criteria inattentive/combined	100	.370	.450	.333	.600	−.058	−.081	−.068	.465
DISC only hyperactive-impulsive/ combined	104	.481	.462	.500	.536	.037 (±.036)	.034 (±.035)	.036 (.014 − .058)	.518
Multiple criteria hyperactive-impulsive/ combined	100	.320	.450	.311	.673	−.013	−.023	−.017	.491
Commission errors									
DISC only all subtypes	104	.798	.144	.733	.191	−.321	−.014	−.026	.471
Multiple criteria all subtypes	100	.440	.140	.429	.558	−.020	−.004	−.007	.497
DISC only inattentive/combined	104	.779	.144	.733	.213	−.206	−.010	−.019	.481
Multiple criteria inattentive/combined	100	.370	.140	.357	.628	−.020	−.006	−.009	.496
DISC only hyperactive-impulsive/combined	104	.481	.144	.533	.528	.101 (±.058)	.018 (±.026)	.031 (.003 − .059)	.515
Multiple criteria hyperactive-impulsive/combined	100	.320	.140	.286	.674	−.050	−.017	−.026	.489

[a]P = Prevalence of a positive identification of problems.

[b]Q = Level of the test (Base rate of participants with a CCPT Overall Index score at or above the cutoff).

[c]PPP = Positive Predictive Power: proportion of participants who test positive for ADHD symptoms on the CCPT Overall Index who were classified as having a problem.

[d]NPP = Negative Predictive Power: proportion of participants who test negative for ADHD symptoms on the CCPT Overall Index who were classified as having no problem.

[e]cPPP = PPP recalibrated to correct for prevalence, yields score that ranges from 0 (random test) to 1 (perfect test).

[f] 95% confidence interval.

[g]cNPP: NPP recalibrated to correct for prevalence; yields score that ranges from 0 (random test) to 1 (perfect test).

[h]κ = Kappa statistic; Quality (recalibration) of efficiency. Efficiency is defined as the probability that classification based on the CCPT and classification based on other measures of ADHD behaviors agree (True Positive Rate + True Negative Rate; 0 = no agreement other than that which would be expected by chance; 1.00 = total agreement).

[i]AUC = Area under the ROC curve: Index of identification accuracy (.50 = no better than chance prediction; 1.00 = perfect rate of prediction).

[j]Identified group = positive DISC any subtype & positive parent and teacher CADS; Non-identified group = negative DISC or subthreshold cases.

[k]Subthreshold and hyperactive-impulsive cases included in non-identified group.

[l]Subthreshold and inattentive only cases included in non-identified group.

These findings also suggest differences between subtypes of ADHD, which is important given the current debate about whether the inattentive subtype is a qualitatively distinct disorder from the hyperactive-impulsive and combined subtypes.

This study has several limitations which should be considered and can inform future research. The generalizability of this study is limited by the approaches used to classify problems with inattentive and hyperactive-impulsive behaviors. A positive parent diagnostic interview by itself is insufficient for diagnosing ADHD, according to current standards of practice. The addition of parent and teacher behavioral ratings provided more conservative criteria for classification and, in most cases, improved accuracy over the interview alone; however, current diagnostic standards and decision-making for ADHD require consideration of additional assessment methods and criteria, such as the assessment of functional impairment and consideration of comorbidities and exclusions. There are several factors which present a challenge for establishing criteria for classification, even with the current standards. These include the dimensional nature of ADHD symptoms, different subtypes of ADHD, and multiple methods and informants. These factors require the establishment of cutoffs and a method of integrating results from different methods and settings. How these are handled will affect identification rates and the ability to compare results across studies. The results of this study suggest that more conservative criteria will result in improved accuracy for the CCPT. Future studies should examine the correspondence between the CCPT and different algorithms for an ADHD gold standard that align with the current standards. Lastly, given the differing parameters and versions of the CPT, this study cannot be easily generalized to the other versions of the CPT that are marketed and in use. Additional studies can compare the diagnostic precision of the different CPTs.

An objective measure of the core symptoms of ADHD with sufficient validity and accuracy would be a welcome addition to ADHD assessment armamentarium. Since children with ADHD are a heterogeneous group, different objective tests will need to be matched with the different subtypes. This study found the CCPT to be more accurate in identifying problems with hyperactive-impulsive behaviors, which is consistent with current thinking that these behaviors are the result of an underlying problem with behavioral inhibition (Barkley, 1997). However, the accuracy of the CCPT with this subgroup was found to be fair to slight at best. Additional research needs to clarify the underlying constructs that the different objective tests measure, as well as establish their accuracy in identifying the different subtypes of ADHD. Until that time, obtaining information from informants who know the child well via reliable methods remains the standard and clinicians should be cautious in how CCPT results are integrated into these other standard evaluation methods.

References

Achenbach, T. M. (1991a). *Manual for the child behavior checklist/4-18 and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the teacher report form and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

American Academy of Child and Adolescent Psychiatry. (1991). Practice parameters for the assessment and treatment of attention-deficit hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*(3), I–III.

American Board of Pediatrics. (2000). Clinical practice guideline: Diagnosis and evaluation of the child with attention-deficit/hyperactivity disorder. *Pediatrics*, *105*(5), 1158–1170.

American Psychiatric Association. (1994). *Diagnostic and statistical manual for mental disorders* (4th edn.). Washington, DC: Author.

Aylward, G. P., Verhulst, S. J., & Bell, S. (1990). Individual and combined effects of attention deficit and learning disabilities on computerized ADHD assessment. *Journal of Psychoeducational Assessment*, *8*, 497–508.

Ballard, J. C. (1996). Computerized assessment of sustained attention: Interactive effects of task demand, noise, and anxiety. *Journal of Clinical and Experimental Neuropsychology*, *18*(6), 864–882.

Barkley, R. A. (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *Journal of Abnormal Child Psychology*, *19*(2), 149–178.

Barkley, R. A. (1997). *ADHD and the nature of self-control*. New York: Guilford Press

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

Cohen, J. (1988). *Statistical power and analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Columbia University DISC Development Group. (2000). *C-DISC 4 for windows: User manual*.

Conners, C. K. (1994). *Conner's continuous performance test*. Toronto, Canada: Multi-Health Systems.

Corkum, P. V., & Siegel, L. S. (1993). Is the continuous performance task a valuable research tool for use with children with attention-deficit-hyperactivity disorder? *Journal of Child Psychology and Psychiatry*, *34*, 1217–1239.

Conners, C. K. (1994). *The Conners continuous performance test*. Toronto, Canada: Multi-Health Systems.

Conners, C. K. (1997). *Conners' rating scales*. Toronto, Canada: Multi-Health Systems.

DuPaul, G. J., Anastopoulus, A. D., Shelton, T. L., Guevremont, D. C., & Metevia, L. (1992). Multimethod assessment of attention deficit hyperactivity disorder: The diagnostic utility of clinic based tests. *Journal of Clinical Child Psychology*, *21*(4), 394–402.

Dupuy, T. R., & Greenberg, L. M. (1993). *TOVA manual*. Minneaplois, MN: Lawrence M. Greenberg.

Epstein, J. N., Conners, C. K., Sitarenios, G., & Erhardt, D. (1998). Continuous performance test results of adults with attention deficit hyperactivity disorder. *The Clinical Neuropsychologist*, *12*(2), 155–168.

Epstein, J. N., Erkanli, A., Conners, C. K., Klaric, J., Costello, J. E., & Angold, A. (2003). Relations between continuous performance test performance measures and ADHD behaviors. *Journal of Abnormal Child Psychology*, *31*(5), 543–554.

Forbes, G. B. (1998). Clinical utility of the test of variables of attention (TOVA) in the diagnosis of attention-deficit/hyperactivity disorder. *Journal of Clinical Psychology*, *54*(4), 461–476.

Furlong, M. J. (1998). Review of the child behavior checklist. In The thirteenth mental measurements yearbook. Retrieved October 27, 2006, from EBSCOHost Mental Measurements Yearbook database.

Gadow, K. D., & Sprafkin, J. (1997). *Child symptom inventory-4*. Stony Brook, NY: Checkmate Plus.

Gordon, M. (1986). *Instruction manual for the Gordon diagnostic system*. DeWitt, NY: Gordon Diagnostic Systems.

Halperin, J. M., Matter, K., Bedi, G., Sharma,V., & Newcorn, J. H. (1992*)*. Specificity of inattention, impulsivity, and hyperactivity to the diagnosis of attention-deficit hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *31*, 429–436.

Halperin, J. M., Sharma, V., Greenblatt, E., & Schwartz, S. (1991). Assessment of the continuous performance test: Reliability and validity in the nonreferred sample. *Psychological Assessment*, *3*(4), 603–608.

Halperin, J. M., Wolf, L. E., Greenblatt, E., & Young, G. (1991). Subtype analysis of commission errors on the continuous performance test in children. *Developmental Neuropsychology*, *7*(2), 207–217.

Jensen, P., & Watanabe, H. (1999). Sherlock Holmes and child psychopathology assessment approaches: The case of the false-positive. *Journal of Child and Adolescent Psychiatry*, *38*(2), 138–146.

Kaufman, A. D., & Kaufman, N. L. (1990). *Kaufman brief intelligence test manual (KBIT)*. Circle Pines, MN: American Guidance Service.

Klee, S. H., & Garfinkel, B. D. (1983). The computerized performance task: a new measure of inattention. *Journal of Abnormal Child Psychology*, *11*, 487–495.

Koelega, H. S. (1995). Is the continuous performance task useful in research with ADHD children? Comment on a review. *Journal of Child Psychology and Psychiatry*, *36*(8), 1477–1485.

Kraemer, H. C. (1992). *Evaluating medical tests*. Thousand Oaks: Sage Publications.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Losier, B. J., McGrath, P. J., & Klein, R. M. (1996). Error patterns on the Continuous Performance Test in non-medicated and medicated samples of children with and without ADHD: A meta-analytic review. *Journal of Child Psychology and Psychiatry*, *37*(8), 971–987.

McGee, R. A., Clark, S. E., & Symons, D. K. (2000). Does the Conners' continuous performance test aid in ADHD diagnosis? *Journal of Abnormal Child Psychology*, *28*(5), 415–424.

Nichols, S. L., & Waschbusch, D. A. (2004). A review of the validity of laboratory cognitive tasks used to assess symptoms of ADHD. *Child Psychiatry and Human Development*, *34*(4), 297–315.

O'Brien, J. D., Halperin, J. M., Newcorn, J. H., Sharma, V., Wolf, L., & Morganstein, A. (1992). Psychometric differentiation of conduct disorder and attention deficit disorder with hyperactivity. *Journal*

of the American Academy of Child and Adolescent Psychiatry, *24*, 429–436.

O'Dougherty, M., Nuechterlein, K. H., & Drew, B. (1984). Hyperactive and hypoxic children: Signal detection, sustained attention, and behavior. *Journal of Abnormal Psychology*, *93*, 178–191.

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 Psychology*, *34*(3), 449–476.

Pfiffner, L. J., McBurnett, K., Lahey, B. B., Loeber, R., Green, S., Frick, P. J., & Rathouz, P. J. (1999). Association of parental psychopathology to the *comorbid disorders* of boys with *attention deficit-hyperactivity disorder*. *Journal of Consulting and Clinical Psychology*, *67*(6), 881–893.

Rapport, M. D., Chung, K., Shore, G., Denney, C. B., & Isaacs, P. (2000). Upgrading the science and technology of assessment and diagnosis: Laboratory and clinic-based assessment of children with ADHD. *Journal of Clinical Child Psychology*, *29*(4), 555–568.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., Jr., & Beck L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, *20*(5), 343–350.

Rovet, J. F., & Hepworth, S. L. (2001). Dissociating attention deficits in children with ADHD and congenital hypothyroidism using multiple CPTs. *Journal of Child Psychology and Psychiatry*, *42*(8), 1049–1056.

Schatz, A. M., Ballantyne, A. O., & Trauner, D. A. (2001). Sensitivity and specificity of a computerized test of attention in the diagnosis of attention-deficit/hyperactivity disorder. *Assessment*, *8*(4), 357–365.

Shapiro, S. K., & Garfinkel, B. D. (1986). The occurrence of behavior disorders in children: The interdependence of attention deficit and conduct disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *25*, 809–819.

Szatmari, P., Offord, D. R., & Boyle, M. H. (1989). Ontario Child Health Study: Prevalence of attention deficit disorder with hyperactivity. *Journal of Consulting and Clinical Psychology*, *30*, 219–230.

Trommer, B. L., Hoeppner, J. B., Lorber, R., & Armstrong, K. (1988). Pitfalls in the use of a continuous performance test as a diagnostic tool in attention deficit disorder. *Developmental and Behavioral Pediatrics*, *9*(6), 339–345.

Vance, B., & Fuller, G. B. (1995). Relation of scores on WISC-III and WRAT-3 for a sample of referred children and youth. *Psychological Reports*, *76*(2), 371–374.

Wilens T. E., Biederman, J., Brown, S., Tanquay, S., Monuteaux, M. C., Blake, C., & Spencer, T. J. (2002). Psychiatric comorbidity and functioning in clinically referred preschool children and school-age youths with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(3), 262–268.

Wilkinson, G. S. (1993). *Wide range achievement test* (3rd ed.). Wilmington, DE: Wide Range.

Zahn, T. P., Kruesi, M. J. P., & Rapoport, J. L. (1991). Reaction time indices of attention deficits in boys with disruptive disorders. *Journal of Abnormal Child Psychology*, *19*, 233–252.