

Standardized Observational Assessment of Attention Deficit Hyperactivity Disorder Combined and Predominantly Inattentive Subtypes. I. Test Session Observations

Stephanie H. McConaughy and Masha Y. Ivanova
University of Vermont

Kevin Antshel
SUNY Upstate Medical University

Ricardo B. Eiraldi
University of Pennsylvania

Abstract. Test examiners used the Test Observation Form (McConaughy & Achenbach, 2004) to rate test session behavior of 177 6- to 11-year-old children during administration of the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV) and Wechsler Individual Achievement Tests—Second Edition (WIAT-II). Participants were assigned to four groups based on a parent diagnostic interview and parent and teacher rating scales: attention deficit hyperactivity disorder (ADHD)—Combined type ($n = 74$); ADHD—Inattentive type ($n = 25$); clinically referred without ADHD ($n = 52$); and controls ($n = 26$). The ADHD—Combined type group scored significantly higher than the other three groups on six Test Observation Form scales: (1) Attention Problems; (2) Oppositional; (3) Attention Deficit/Hyperactivity Problems scale; (4) Inattention subscale; (5) Hyperactivity-Impulsivity subscale; and (6) Externalizing. The two ADHD groups also scored significantly lower than controls on all WISC-IV and WIAT-II composites and lower than those clinically referred without ADHD on WISC-IV Working Memory Index and Full Scale Intelligence Quotient. Implications are discussed regarding the discriminative validity of standardized test session observations for identifying children with ADHD and differentiating between the two ADHD subtypes.

Preparation of this article was supported in part by Grant R01 HD40220 from the National Institute of Child Health and Human Development to Stephanie McConaughy at the University of Vermont and by the University of Vermont Research Center for Children, Youth, and Families. Statements do not represent the position or policy of these agencies and no official endorsement by them should be inferred. The first author has a potential conflict of interest because she is the developer of the Test Observation Form, the primary measure used in this study. However, the second author managed all final data sets and conducted the majority of analyses for this study.

Correspondence regarding this article should be addressed to Stephanie H. McConaughy, Department of Psychiatry, University of Vermont, One South Prospect Street, Burlington, VT 05401; E-mail: stephanie.mcconaughey@uvm.edu

Copyright 2009 by the National Association of School Psychologists, ISSN 0279-6015, which has nonexclusive ownership in accordance with Division G, Title II, section 518 of P. L. Law 110-161 and NIH Public Access Policy.

The *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition—Text Revision (DSM-IV-TR)*; American Psychiatric Association, 2000) defines three subtypes of attention deficit hyperactivity disorder (ADHD): Predominantly Inattentive (hereafter, ADHD-IN), showing 6 of 9 symptoms of inattention, but fewer than 6 symptoms of hyperactivity and impulsivity; Predominantly Hyperactive-Impulsive, showing the opposite symptom pattern; and Combined (hereafter, ADHD-C), showing at least 6 of 9 symptoms of inattention and 6 of 9 symptoms of hyperactivity and impulsivity. The *DSM-IV-TR* also specifies that the ADHD symptoms must have persisted for at least 6 months to a degree that is developmentally deviant; at least some of the symptoms were present before age 7; the symptoms cause functional impairment in two or more settings (e.g., at home and at school or work); and there is evidence of clinically significant functional impairment in social, academic, or occupational functioning. Consistent with the *DSM-IV-TR*, Barkley (1997, 2006) has also postulated that ADHD-C is characterized by a core deficit in inhibitory control, which is manifested in poor behavioral inhibition and hyperactivity, combined with inattention and disorganization, whereas ADHD-IN is characterized by inattention and disorganization, without a core deficit in inhibitory control.

Many studies have shown that parent and teacher ratings of attention problems, hyperactivity, and impulsivity distinguish children with ADHD from those without ADHD (for reviews, see Barkley, 2006; Brock & Clinton, 2007; DuPaul & Stoner, 2003; Nigg, 2006). However, Barkley (1997) has argued that relying on parent and teacher ratings in validity studies of ADHD is circular because parent and teacher reports were the primary data used to create the diagnostic criteria in the first place. To avoid circularity, Barkley maintained that external validators are needed in addition to parent and teacher ratings. Similarly, the National Institutes of Health consensus panel (National Institutes of Health, 2000) concluded that although ADHD diagnoses can be made reliably with structured parent diagnostic interviews, there is still no indepen-

dently valid diagnostic or laboratory test for ADHD. Moreover, studies of neuropsychological and laboratory tests have shown few, if any, significant differences between the ADHD-C and ADHD-IN subtypes (Chhabildas, Pennington, & Wilcutt, 2001; Nigg, Blaskey, Huang-Pollock, & Rappley, 2002; Solanto et al., 2007).

Test Session Observations

In the absence of definitive neuropsychological or laboratory tests for ADHD, direct observations of children's behavior by *independent observers* may be one avenue for obtaining external validation of ADHD diagnoses. In fact, many experts consider systematic direct observations of children's behavior to be essential components of clinical assessment (e.g., Mash & Terdal, 2000; Sattler, 2008; Shapiro & Kratochwill, 2000). Test sessions offer fertile arenas for directly observing behavioral manifestations of poor behavioral inhibition, hyperactivity, inattention, and disorganization that have been considered core features of ADHD. As Glutting and colleagues noted, "None of the major contexts of child development (e.g., home, school, and community) offers as high a level of professional expertise, observational control, or uniformity of conditions as the context of individual test-taking" (Glutting, Youngstrom, Oakland, & Watkins, 1996, p. 94).

Compared to other observational settings, test sessions have several advantages. First, tests are usually conducted under standardized conditions that are less variable than conditions in other settings at home or in school classrooms. Second, test examiners can compare their observations of an individual child to observations of other children under similar conditions. Third, test examiners can use their observations to evaluate the validity of test scores for a given child. Fourth, test session observations have the advantage of being relatively "objective" when obtained by test examiners who have no special relationship with a child, especially if test examiners are "blinded" to referral complaints. Thus, although test examiners may develop some hy-

potheses about a child during testing, their observations can still add a somewhat “independent” perspective in contrast to parent and teacher reports (McConaughy, 2005a,b).

Although parent and teacher reports still form the primary basis for diagnosing ADHD, intelligence and achievement tests are often conducted for differential diagnosis of the disorder and to determine special education eligibility. For example, Demaray, Schaefer, and Delong (2003) reported that 73.1% of school psychologists administered intelligence tests, and 67.4% administered achievement tests as components of an ADHD assessment. Reid, Maag, and Vasa (1993) also reported that approximately 50% of children with ADHD qualify for special education services, which usually requires cognitive and achievement testing. Cognitive and achievement test results can help to rule in or rule out other disorders (e.g., mental retardation, learning disability) that may better account for a referred child’s inattention and/or behavioral problems. Cognitive and achievement test results can also help to identify academic deficits that may warrant intervention along with interventions targeting ADHD symptoms (DuPaul, 2007; DuPaul & Stoner, 2003).

Two previous studies reported differences in test session behavior for children with ADHD versus typically developing control children. Based on informal observations of test session behavior, Teicher, Ito, Glod, and Barber (1996) found that children with ADHD moved more frequently than control children during a continuous performance test (CPT). In a more rigorous study based on observations during the Wechsler Intelligence Test—Third Edition (WISC-III; Wechsler, 1991), Glutting, Robins, and de Lancy (1997) found that children with ADHD scored significantly higher than controls on three scales from the Guide to Assessment of Test Session Behavior (Glutting & Oakland, 1993): Inattentiveness, Uncooperative Mood, and Avoidance. The Guide to Assessment of Test Session Behavior Inattentiveness scale contributed most to discrimination between the two groups.

Two additional studies examined test session observations for clinical samples of

children with ADHD-consistent problems versus clinically referred children without ADHD. Gordon, DiNiro, Mettelman, and Tallmadge (1989) tested relations between scores on a CPT, teacher ratings of children’s problems, and test examiners’ ratings of observed behavior during the CPT. Test examiners used a five-point Likert scale to rate the degree to which 6- to 12-year-old children looked away from the CPT, verbalized during the test, and got out of their seats. Gordon et al. reported moderate correlations of .35 between test examiners’ ratings and CPT scores and .25 between test examiners’ ratings and teacher ratings of attention problems (which included hyperactivity). For broad indices of abnormal versus normal attention problems, test examiners’ ratings agreed with CPT scores or teacher ratings for 70% of cases. Agreement across all three measures of attention problems occurred for 50% of cases.

Willcutt, Hartung, Lahey, Loney, and Pelham (1999) examined the diagnostic utility of test session observations for preschool children with *DSM-IV* (American Psychiatric Association, 1994) diagnoses of ADHD versus clinically referred children without ADHD. Test examiners rated children on three items (motor activity, distractibility, and impulse control) of the Hillside Behavior Rating Scale (Gittelman & Klein, 1985) after administering the short form of the Stanford Binet Intelligence Scale—Fourth Edition (Thorndike, Hagan, & Sattler, 1986) and two standardized achievement tests. Willcutt et al. reported correlations of .32 to .50 between the test examiners’ Hillside Behavior Rating Scale scores and the number of ADHD symptoms reported by parents and teachers, as well as significant correlations between the test examiners’ Hillside Behavior Rating Scale scores and several measures of social impairment. Although parent and teacher reports accounted for a large proportion of the variance in intelligence and achievement test scores and social impairment, test examiners’ ratings on the Hillside Behavior Rating Scale items added significant unique variance for predicting impairment over and above parent and teacher reports. However, like the Gordon et al. (1989) study,

the Willcutt et al. (1999) study was limited in its focus on only a few broad descriptors of ADHD-consistent test session behaviors.

We know of only two previous studies that examined test session observations for the ADHD subtypes or their equivalents. Barkley, DuPaul, and McMurray (1990) found that children with *DSM-III* (American Psychiatric Association, 1980) diagnoses of attention deficit disorder with hyperactivity exhibited more ADHD-consistent behaviors during testing (e.g., being off-task, fidgeting, and out-of-seat) than did children with attention deficit disorder without hyperactivity. In a more recent study, Solanto et al. (2007) used a battery of 11 neurocognitive tests, plus the WISC-III and Wechsler Individual Achievement Test (WIAT; Wechsler, 1992), to test differences between children with *DSM-IV* diagnoses of ADHD-IN and ADHD-C and typically developing controls. Their measures included a 10-item checklist on which test examiners rated children's off-task behavior on a three-point scale during administration of a CPT. Solanto et al. (2007) found significant differences between the two ADHD groups versus controls and between ADHD-C and ADHD-IN subtypes on several neuropsychological measures, plus WISC-III and WIAT-II test scores. However, all but two of the significant group effects on test scores disappeared when WISC-III Full Scale Intelligence Quotient (FSIQ) was added as a covariate in analyses. By contrast, test examiners' ratings of off-task behavior produced significantly higher scores for children with ADHD-C versus ADHD-IN even when FSIQ was included as a covariate. A limitation of the Solanto et al. study, however, was that test examiners were not blind to children's diagnostic group assignment.

Purpose of the Present Study

In the present study, we obtained test examiners' ratings of test session behaviors for children with *DSM-IV-TR* diagnoses of ADHD-C and ADHD-IN, plus clinically referred children without ADHD (NON-ADHD-REF) and typically developing nonreferred controls (Control). Children in all four groups

were administered the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003) and Wechsler Individual Achievement Tests—Second Edition (WIAT-II; Wechsler, 2002). Unlike previous studies, we used a much more comprehensive rating form, the Test Observation Form (TOF; McConaughy & Achenbach, 2004), on which test examiners rate an array of test session behaviors, including behaviors consistent with symptoms of ADHD, internalizing and externalizing problems, and test-taking behaviors.

The TOF is a standardized rating form developed as part of the Achenbach System of Empirically Based Assessment (Achenbach & Rescorla, 2001). The TOF scoring profile provides summative raw scores, standard scores, and percentiles for five empirically based syndrome scales: Internalizing, Externalizing and Total Problems; and a *DSM*-oriented Attention Deficit/Hyperactivity Problems (ADHP) scale and Inattention and Hyperactivity-Impulsivity subscales (for details, see Methods section).

In the present study, we had the following hypotheses: (1) children with ADHD-C would score significantly higher than a non-ADHD referred group and Control on the empirically based TOF Attention Problems syndrome, plus the ADHP scale and ADHP Inattention and Hyperactivity-Impulsivity subscales; (2) children with ADHD-C and ADHD-IN would score significantly higher than NON-ADHD-REF on TOF Attention Problems, the ADHP scale, and ADHP Inattention subscale; and (3) children with ADHD-C would score significantly higher than children with ADHD-IN on TOF Attention Problems and the ADHP scale (both of which include hyperactivity and impulsivity along with inattention), and ADHP Hyperactivity-Impulsivity subscale. Because children with both ADHD subtypes should have problems with inattention, we expected no differences between ADHD-C and ADHD-IN on the ADHP Inattention subscale. Because children with ADHD often have other comorbid diagnoses, we made no a priori hypotheses regarding differences between children with ADHD and NON-ADHD-REF on other TOF problem scales. In addition to testing specific hypothe-

ses regarding the TOF, we also examined group differences on the WISC-IV and WIAT-II. We limited our research to ages 6–11 in order to tap symptoms of ADHD in school-age children before they experienced developmental changes that occur in adolescence.

Method

Participants

This study was part of a larger federally funded research effort to test the contribution of standardized observations of test session and classroom behavior for improving assessment of ADHD. Participants were recruited from mental health providers and public and private schools in the vicinity of outpatient clinics at three study sites: the Vermont Center for Children, Youth, & Families at the University of Vermont, Department of Psychiatry in Burlington, Vermont (UVM; $n = 56$); the Children's Hospital of Philadelphia in Philadelphia, Pennsylvania (CHOP, $n = 54$); and the Child and Adolescent Psychiatry Clinic at SUNY Upstate Medical University in Syracuse, New York (SUNY, $n = 67$). The UVM clinic was in a semirural, small urban area and the CHOP and SUNY clinics were in large urban centers. Participants for the present study were drawn from a total sample of 456, 6- to 11-year-old children participating in the larger study.

Diagnostic Group Assignment

To be assigned to the ADHD-C group, parents and/or teachers had to report symptoms of both inattention and hyperactivity-impulsivity: that is, the child had to have a positive DSM-IV diagnosis of ADHD—Combined type (314.01) on the NIMH Diagnostic Interview Schedule for Children—Fourth Edition (DISC-4; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), plus scores ≥ 80 th percentile on the Inattention or Hyperactivity-Impulsivity subscales of the School version of the ADHD Rating Scale—IV (ADHDRS-IV; DuPaul, Power, Anastopolous, & Reid, 1998; see Measures section for description of instruments). To be assigned to the ADHD-IN group, the child had to have a positive diag-

nosis of ADHD—Inattentive type (314.00) on the NIMH DISC-4, plus a score ≥ 80 th percentile on the Inattention subscale of the ADHDRS-IV—School version and a score < 80 th percentile on the Hyperactivity-Impulsivity subscale of the ADHDRS-IV—School version. To be assigned to the NON-ADHD-REF group, the child had to have no diagnosis of ADHD on the NIMH DISC-4, plus scores < 80 th percentile on the Inattention and Hyperactivity-Impulsivity subscales of the ADHDRS-IV—School version. To be assigned to the Control group, the child had to have no diagnosis of ADHD on the NIMH DISC-4, plus scores < 80 th percentile on the Inattention and Hyperactivity-Impulsivity subscales of the ADHDRS-IV—School version and the ADHDRS-IV—Home version. Additional criteria for recruitment of Control children were the following: at least average estimated cognitive ability, not have repeated a grade, and not have been referred for or received special education, a Section 504 plan, counseling, or mental health services within the past 12 months.

Exclusionary criteria for children in all four groups were as follows: WISC-IV FSIQ < 80 and physical, medical, or mental disabilities that might interfere with learning or test behavior (e.g., seizure disorders, cerebral palsy, mental retardation, autism, or pervasive developmental disorder). For the few children with prescriptions of stimulants or other medications for ADHD, parents agreed not to administer medications on the day the child was tested. Table 1 shows demographic characteristics of the sample.

As shown in Table 1, there were approximately three times as many boys than girls in the ADHD-C and ADHD-IN groups, consistent with rates reported in the *DSM-IV-TR*. Socioeconomic status (SES) was scored on Hollingshead's (1975) nine-point scale based on parental occupation reported by the parent. The mean SES for the total sample was 6.2 ($SD = 1.8$; $n = 167$). A one-way analysis of variance (ANOVA) showed significant group differences on SES ($F[3,163] = 5.18$, $p = .002$), with ADHD-C scoring significantly lower than NON-ADHD-REF ($p = .001$).

Table 1
Demographic Characteristics and DSM-IV Diagnoses for Four Groups

| Characteristics | ADHD-C | ADHD-IN | NON-ADHD-REF | Control |
|---|------------|------------|--------------|------------|
| Boys, <i>n</i> (%) | 57 (77.0%) | 19 (76.0%) | 36 (69.2%) | 15 (57.7%) |
| Girls, <i>n</i> (%) | 17 (23.0%) | 6 (24.0%) | 16 (30.8%) | 11 (42.3%) |
| Total sample, <i>n</i> (%) | 74 (41.8%) | 25 (14.1%) | 52 (29.4%) | 26 (14.7%) |
| Mean age (<i>SD</i>) | 7.5 (1.6) | 8.4 (1.5) | 8.5 (1.7) | 8.6 (1.5) |
| Mean SES (<i>SD</i>) ^a | 5.7 (1.8) | 6.2 (1.8) | 6.9 (1.6) | 6.5 (1.8) |
| Ethnicity, <i>n</i> (%) | | | | |
| Non-Latino White | 41 (55.4%) | 17 (68.0%) | 38 (73.1%) | 14 (53.8%) |
| African American | 26 (35.1%) | 5 (20.0%) | 6 (11.5%) | 8 (30.8%) |
| Other or Unknown | 7 (9.5%) | 3 (12.0%) | 8 (15.4%) | 4 (15.4%) |
| DSM-IV diagnoses, <i>n</i> (%) ^b | | | | |
| ADHD only | 23 (31.1%) | 14 (56.0%) | 0 (0%) | 0 (0%) |
| Conduct disorder | 12 (16.2%) | 1 (4.0%) | 0 (0%) | 1 (3.8%) |
| Dysthymia or major depression | 6 (8.1%) | 2 (8.0%) | 3 (5.8%) | 0 (0%) |
| Generalized anxiety disorder | 5 (6.8%) | 2 (8.0%) | 3 (5.8%) | 0 (0%) |
| Obsessive compulsive disorder | 4 (5.4%) | 1 (4.0%) | 3 (5.8%) | 1 (3.8%) |
| Oppositional defiant disorder | 39 (52.7%) | 6 (24.0%) | 14 (26.9%) | 1 (3.8%) |
| Separation anxiety | 15 (20.3%) | 3 (12.0%) | 5 (9.6%) | 0 (0%) |
| Social phobia | 4 (5.4%) | 2 (8.0%) | 3 (5.8%) | 0 (0%) |
| Specific phobia | 25 (33.8%) | 5 (20.0%) | 10 (19.2%) | 6 (23.1%) |
| Tourette's or tic disorder | 6 (8.1%) | 1 (4.0%) | 4 (7.7%) | 0 (0%) |
| No diagnosis, <i>n</i> (%) | 0 (0%) | 0 (0%) | 25 (48.1%) | 19 (73.1%) |
| One diagnosis, <i>n</i> (%) | 23 (31.1%) | 14 (56.0%) | 16 (30.8%) | 6 (23.0%) |
| Two or more diagnoses, <i>n</i> (%) | 51 (68.9%) | 11 (44.0%) | 11 (21.2%) | 1 (3.8%) |

Note. Total *N* = 177. ADHD-C = attention deficit hyperactivity disorder—combined type; ADHD-IN = ADHD—inattentive type; NON-ADHD-REF = non-ADHD clinically referred; Control = nonreferred controls; *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*.

^a SES scored on Hollingshead's (1975) scale, where 1 = lowest and 9 = highest.

^b Children with comorbid diagnoses were counted more than once for the different diagnostic categories.

There were also site differences in SES ($F[2,164] = 14.50, p < .001$), with CHOP cases showing significantly lower SES (mean = 5.2, $SD = 1.6$) than UVM cases (mean = 6.8, $SD = 1.7$) and SUNY cases (mean = 6.5, $SD = 1.6; p < .001$). CHOP also had higher percentages of African American cases. Ethnicity of the total sample was 61.6% non-Latino White, 26% African American, 3.4% Latino or Hispanic, 6.8% mixed, and 2% other or unknown.

The bottom half of Table 1 shows the number of cases with various *DSM-IV* diagnoses, according to symptom criteria reported

by parents on the DISC-4. Children in the two ADHD groups were allowed to have other comorbid *DSM-IV* diagnoses. Also, children in the NON-ADHD-REF and Control groups were allowed to have diagnoses other than ADHD. As can be seen in Table 1, 31.1% of ADHD-C and 56% of ADHD-IN cases had only ADHD diagnoses, whereas 68.9% of ADHD-C and 44% of ADHD-IN had two or more diagnoses. Within the Control group, 73.1% had no diagnosis. Six of the Control cases had diagnoses of specific phobia, which included fear of the dark and fear of insects.

Measures

ADHDRS-IV. The ADHDRS-IV (DuPaul et al., 1998) is an 18-item rating scale, with 9 items that assess *DSM-IV*-defined symptoms of inattention and 9 items that assess *DSM-IV*-defined symptoms of hyperactivity-impulsivity. Each item is rated on a 4-point scale: 0 = *not at all, rarely*; 1 = *sometimes*; 2 = *often*; and 3 = *very often*. The ADHDRS-IV—Home version is completed by parents, and the ADHDRS-IV—School version is completed by teachers. Raw scores, *T* scores, and percentiles are provided for Total Problems, Inattention, and Hyperactivity-Impulsivity based on large stratified national samples. The three ADHDRS-IV scales showed internal consistencies ranging from .86 to .96. Test–retest reliabilities over a 4-week interval were: Total Problems = .85 and .90, Inattention = .78 and .89, Hyperactivity-Impulsivity = .86 and .88, for the Home and School versions, respectively.

NIMH DISC-4. The NIMH DISC-4 (Shaffer et al., 2000) is a highly structured diagnostic interview administered to parents to assess criteria for *DSM-IV* disorders applicable to children ages 6–17. Diagnoses are assessed for the past 12 months and past 4 weeks. For this study, we administered the computer-assisted NIMH DISC-4 modules for ADHD, conduct disorder, oppositional defiant disorder, anxiety disorders, and mood disorders. Test–retest kappas for the NIMH DISC-4 were .96 for specific phobia, .79 for ADHD, .66 for major depression, .65 for generalized anxiety, .58 for separation anxiety, .54 for oppositional defiant disorder and social phobia, and .43 for conduct disorder.

TOF. The Test Observation Form (TOF; McConaughy & Achenbach, 2004) is a standardized rating form to be completed by test examiners. The TOF contains 125 items that describe children’s behavior, affect, and test-taking style. During test administration, examiners record narrative observations of the child’s behavior in space provided on the TOF or on the test protocol. Immediately after completing the test, examiners rate the child on

each problem item, using a 4-point scale: 0 = *no occurrence*; 1 = *very slight or ambiguous occurrence*; 2 = *definite occurrence with mild to moderate intensity and less than 3 minutes duration*; 3 = *definite occurrence with severe intensity or 3 or more minutes duration*.

The TOF is scored on five syndrome scales (Withdrawn/Depressed, Language/Thought Problems, Anxious, Oppositional, and Attention Problems); Internalizing (consisting of Withdrawn/Depressed and Language/Thought Problems) and Externalizing (consisting of Oppositional and Attention Problems); Total Problems; and *DSM*-oriented Attention Deficit/Hyperactivity Problems (ADHP) scale and Inattention and Hyperactivity-Impulsivity subscales.

McConaughy and Achenbach (2004) reported internal consistencies ranging from .74 to .94 for the 11 TOF scales. Test–retest reliabilities over an average interval of 10 days ranged from .53 to .87 across the 11 TOF scales, with a mean test–retest reliability of .80. Test–retest reliabilities of most interest for the present study were .85 for Attention Problems, .87 for the ADHP scale, .82 for ADHP Inattention, and .85 for ADHP Hyperactivity-Impulsivity. Interrater reliabilities ranged from .42 to .79, with a mean interrater reliability of .62. Interrater reliabilities of most interest for this study were .71 for Attention Problems, .73 for the ADHP scale, .67 for ADHP Inattention, and .73 for ADHP Hyperactivity-Impulsivity. Criterion-related validity was demonstrated by significantly ($p < .05$) higher scores for clinically referred than non-referred 6- to 11-year-old children on all TOF scales. For the present study, all analyses were conducted on TOF raw scale scores to maximize variance.

WISC-IV. The WISC-IV (Wechsler, 2003) is a well-known individually administered intelligence test for ages 6–16 years. It provides a FSIQ, subtest scores, and composite scores for a Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI). Standard scores for FSIQ and the four Index scales have a mean of 100 and standard deviation of 15. The

WISC-IV manual reports high internal consistencies and test–retest reliabilities for the Index scales and FSIQ.

WIAT-II. The WIAT-II (Wechsler, 2002) is an individually administered test of academic achievement for ages 4–85 years. It provides a total score plus subtest scores and four composite scores: Reading Composite (RC), Mathematics Composite (MC), Written Language Composite (WLC), and Oral Language Composite. Standard scores for each composite have a mean of 100 and standard deviation of 15. The WIAT-II manual reports high internal consistencies and test–retest reliabilities for the composite scores. For this study, we used RC, MC, and WLC scores.

Procedure

Recruitment of participants. The research protocol was approved by the institutional review boards of each of the three sites. To recruit participants for the three clinically referred groups (ADHD-C, ADHD-IN, NON-ADHD-REF), researchers provided mental health clinicians and school personnel packets of letters and consent forms to give to parents that described the goals and procedures of the study. In order not to bias selection toward concerns about ADHD per se, letters to parents described the study as an effort “to develop procedures for observing children’s behavior in their classrooms and during cognitive testing.” Parents were informed that the study required approximately 3 hr of individual testing with their child (usually done in one test session), plus completion of rating scales by parents and teachers, and observations of their child in test sessions and in school classrooms. Parents mailed consent forms directly to the research staff. Once a parental consent form was received, the researchers arranged appointments for testing the child at the clinic and contacted school staff to arrange observations of the child in the classroom. Parents of referred children were paid \$15 for their participation.

To recruit Control children, researchers provided school personnel packets of study materials containing written instructions and

exclusionary criteria for selecting typically developing children. School staff randomly selected three boys and two girls from each participating classroom. A school staff member mailed letters and consent forms to parents of eligible children. After parents returned signed consent forms to the research staff, appointments were arranged for testing and observations of the child, as done for the referred children. Parents of Control children were paid \$50 for their participation.

Researchers mailed or hand delivered letters to teachers, along with consent forms signed by parents and rating scales to be completed by the teacher for each child. Teachers were informed that the student was “participating in a study of children’s behavioral development.” Teachers were kept blinded to the diagnostic group status of the child throughout the study. Teachers signed consent forms and were paid \$15 for their participation.

Test administration. Test examiners were upper level psychology or school psychology graduate students who had been trained in administration of the WISC-IV as part of their graduate program. Researchers provided additional training in administration of the WISC-IV and WIAT-II and supervised test scoring and interpretation. Administration of the WISC-IV and WIAT-II, along with a CPT, was counterbalanced across child participants. Most children were tested in one session, lasting approximately 3 hr. Children were provided a 15-min break with a sugar-free snack between each test. Test examiners had the option of dividing testing into two sessions on different days for the few children who needed shorter sessions.

TOF ratings. Researchers trained all test examiners in the scoring procedures for the TOF, as described in the manual (McConaughy & Achenbach, 2004). Each test examiner was provided written guidelines for scoring TOF items. For data collection, test examiners recorded notes about a child’s behavior as they administered each test. During the break after each test (and before scoring the test itself), the examiner rated the child on the

125 TOF items. Examiners completed a separate TOF for the WISC-IV, WIAT-II, and CPT. Prior to testing, test examiners were kept blinded to all referral and background information about the child, results of parent and teacher rating scales and interviews, and the child's diagnostic group assignment for the study.

Results

Data Analyses

To test differences on the TOF scales, we performed a series of 2×4 multivariate analyses of variance (MANOVA), treating gender and diagnostic group (ADHD-C, ADHD-IN, NON-ADHD-REF, and Control) as between-subject variables and raw scores on the following sets of TOF scales as dependent variables: five syndrome scales; Internalizing and Externalizing; and ADHP Inattention and Hyperactivity-Impulsivity subscales (SPSS 15.1 general linear model). Because we included gender as a between-subject variable, we were able to analyze raw scores (not T scores) as dependent variables to maximize variance. Each MANOVA was followed by ANOVAs and post hoc Tukey honestly significant difference tests (Tukey HSD) to examine group differences. The Tukey HSD tests were performed on homogeneous subsets to adjust for unequal sample sizes.¹ We also performed two 2×4 ANOVAs on summative raw scores for TOF Total Problems and the ADHP scale, followed by Tukey HSD tests. We examined effect sizes (ES) indicated by partial η^2 , which can be translated directly into percentage of variance accounted for. According to Cohen's (1988) criteria, ES accounting for 1% to 5.8% of variance are small; 5.9% to 13.7% of variance are medium; and >13.8% of variance are large.

To test differences on WISC-IV and WIAT-II composite scores, we performed two separate 2×4 General Linear Model MANOVAs, treating gender and diagnostic group as between-subject variables and standard scores on the following scales as dependent variables: WISC-IV VCI, PRI, WMI, and PSI; and WIAT-II RC, MC, and WLC. Each

of the two MANOVAs was followed by univariate ANOVAs and post hoc Tukey HSD tests to examine group differences. We also performed a 2×4 ANOVA on WISC-IV FSIQ followed by Tukey HSD tests.

We performed discriminant analyses to determine which combinations of the TOF scales and WISC-IV Index scores or WIAT-II composite scores contributed to discriminating between the following groups: (a) ADHD-C versus NON-ADHD-REF; (b) ADHD-C versus Control; (c) ADHD-C versus ADHD-IN; (d) ADHD-IN versus NON-ADHD-REF; and (e) ADHD-IN versus Control. For each group classification, we treated the following variables as sets of candidate predictors: (a) five TOF syndromes; (b) ADHP Inattention and Hyperactivity-Impulsivity subscales; (c) ADHP scale total score plus WISC-IV VCI, PRI, WMI, and PSI, or WIAT-II RC, MC, and WLC; and (d) significant TOF syndromes identified from (a) or the MANOVAs plus WISC-IV VCI, PRI, WMI, and PSI or WIAT-II RC, MC, and WLC. We used forward stepwise discriminant analyses with $p = .05$ as the maximum entry criterion and $p = .10$ as the minimum removal criterion. The standardized canonical discriminant function coefficients indicate the relative importance of each discriminating variable for predicting group membership (regardless of sign). Cross-validated classification rates were obtained in each discriminant analysis.

Group Differences on TOF and WISC-IV

TOF. The top part of Table 2 shows means and standard deviations on the TOF scales for the four diagnostic groups based on examiners' ratings of children's behavior during the WISC-IV. The overall MANOVA for the five TOF syndromes showed a significant main effect of group ($F[15, 456] = 2.98; p < .001$; partial $\eta^2 = .08$). Subsequent ANOVAs revealed significant group effects ($p < .001$) only for the Oppositional and Attention Problems syndromes, with medium to large ES (partial $\eta^2 = .13$ and $.17$, respectively). The 2×4 ANOVA for the ADHP scale total score revealed a significant main effect of group,

Table 2
Group Differences on TOF and WISC-IV Scales

| Dependent Measures | ADHD-C (<i>n</i> = 74) | ADHD-IN (<i>n</i> = 25) | NON-ADHD-REF (<i>n</i> = 52) | Control (<i>n</i> = 26) | <i>F</i> | <i>p</i> | η^2 |
|----------------------------|----------------------------|------------------------------|----------------------------------|-----------------------------|---------------------------|-----------|----------|
| TOF scales | | | | | | | |
| Withdrawn/Depressed | 0.96 (2.37) | 1.48 (3.14) | 1.88 (3.83) | 1.77 (2.90) | <i>ns</i> | <i>ns</i> | .02 |
| Language/Thought Problems | 1.35 (1.90) | 1.60 (2.33) | 1.48 (2.81) | 0.65 (0.98) | <i>ns</i> | <i>ns</i> | .01 |
| Anxious | 3.53 (3.39) | 2.92 (2.99) | 4.13 (4.41) | 2.77 (2.64) | <i>ns</i> | <i>ns</i> | .02 |
| Oppositional | 5.01 (5.68) _a | 2.36 (2.78) _b | 1.54 (3.05) _b | 0.58 (1.82) _b | <i>F</i> (3, 169) = 8.26 | < .001 | .13 |
| Attention Problems | 12.23 (8.73) _a | 6.88 (5.73) _b | 5.98 (5.51) _b | 4.08 (4.37) _b | <i>F</i> (3, 169) = 11.56 | < .001 | .17 |
| ADHP total score | 15.08 (11.06) _a | 8.20 (6.88) _b | 6.79 (6.69) _b | 4.58 (5.15) _b | <i>F</i> (3, 169) = 12.40 | < .001 | .18 |
| ADHP Inattention | 5.39 (5.43) _a | 2.24 (3.70) _b | 2.19 (2.93) _b | 1.12 (1.77) _b | <i>F</i> (3, 169) = 9.65 | < .001 | .15 |
| ADHP Hyperactivity- | | | | | | | |
| Impulsivity | 9.96 (6.44) _a | 5.96 (4.61) _b | 4.60 (4.41) _b | 3.46 (4.32) _b | <i>F</i> (3, 169) = 10.88 | < .001 | .16 |
| Internalizing | 2.31 (3.12) | 3.08 (4.27) | 3.37 (4.73) | 2.42 (3.22) | <i>ns</i> | <i>ns</i> | .01 |
| Externalizing | 17.24 (13.56) _a | 9.24 (7.30) _b | 7.52 (7.53) _b | 4.65 (5.40) _b | <i>F</i> (3, 169) = 12.11 | < .001 | .18 |
| Total Problems | 34.91 (22.97) _a | 23.28 (14.17) _{a,b} | 23.50 (17.50) _{a,b} | 14.92 (10.99) _b | <i>F</i> (3, 169) = 7.07 | < .001 | .11 |
| WISC-IV scales | | | | | | | |
| VCI | 97.18 (12.39) _a | 94.72 (8.76) _a | 106.73 (14.37) _b | 112.38 (12.75) _b | <i>F</i> (3, 169) = 12.13 | < .001 | .18 |
| PRI | 99.22 (13.84) _a | 98.40 (9.72) _a | 103.15 (13.28) _a | 112.65 (14.47) _b | <i>F</i> (3, 169) = 7.52 | < .001 | .12 |
| WMI | 95.77 (10.20) _a | 92.96 (9.60) _a | 102.46 (11.43) _b | 109.42 (11.06) _c | <i>F</i> (3, 169) = 13.86 | < .001 | .20 |
| PSI | 92.61 (14.58) _a | 91.84 (13.68) _a | 95.85 (11.91) _{a,b} | 103.35 (11.67) _b | <i>F</i> (3, 169) = 4.73 | < .01 | .08 |
| FSIQ | 95.57 (10.95) _a | 93.68 (7.78) _a | 103.54 (13.11) _b | 113.31 (11.53) _c | <i>F</i> (3, 169) = 18.60 | < .001 | .25 |

Note. ADHD-C = attention deficit hyperactivity disorder—combined type; ADHD-IN = ADHD—inattentive type; NON-ADHD-REF = non-ADHD clinically referred; Control = controls; ADHP = Attention Deficit/Hyperactivity Problems; TOF = Test Observation Form; WISC-IV = Wechsler Intelligence Scale for Children—Fourth Edition; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index; FSIQ = Full Scale IQ. Total *N* = 177. Mean (*SD*). TOF scores are raw scores; WISC-IV scores are standard scores. Means in the same row with the same subscripts do not differ significantly at *p* < .05 by Tukey HSD comparisons of homogeneous subsets.

with a large ES (partial $\eta^2 = .18$). The 2×4 MANOVA for the ADHP Inattention and Hyperactivity-Impulsivity subscales showed a significant main effect of group ($F[6, 336] = 6.10; p < .001$; partial $\eta^2 = .10$), and subsequent ANOVAs revealed significant group effects ($p < .001$) for both subscales, with medium to large ES (partial $\eta^2 = .15$ and $.16$, respectively). The 2×4 MANOVA for Internalizing and Externalizing showed a significant main effect of group ($F[6, 336] = 5.96; p < .001$; partial $\eta^2 = .10$), and subsequent ANOVAs revealed significant group effects ($p < .001$) only for Externalizing, with a large ES (partial $\eta^2 = .18$). Finally, the 2×4 ANOVA for Total Problems revealed a significant main effect of group, with a medium ES (partial $\eta^2 = .11$).

Tukey HSD tests showed that the ADHD-C group scored significantly higher than ADHD-IN, NON-ADHD-REF and Control on TOF Oppositional, Attention Problems, the ADHP scale total score, ADHP Inattention, ADHP Hyperactivity-Impulsivity, and Externalizing ($p < .05$). On TOF Total Problems, ADHD-C scored significantly higher than Control, but not higher than ADHD-IN or NON-ADHD-REF. There were no significant differences between ADHD-IN versus NON-ADHD-REF and Control groups on any TOF scale. There were no significant effects of gender or Gender \times Group.

WISC-IV. The bottom part of Table 2 shows means and standard deviations for the four diagnostic groups on the WISC-IV Index scales and FSIQ. The overall MANOVA for the four WISC-IV Index scales showed a significant main effect of group ($F[12, 439] = 5.15; p < .001$; partial $\eta^2 = .11$). Subsequent ANOVAs revealed significant group effects ($p < .001$) for all four Index scores, with medium to large ES (partial $\eta^2 = .08$ to $.20$). The 2×4 ANOVA for WISC-IV FSIQ revealed a significant main effect of group, with a large ES (partial $\eta^2 = .25$).

Tukey HSD tests showed that both ADHD-C and ADHD-IN scored significantly lower than Control on all five WISC-IV scales and lower than NON-ADHD-REF on

WISC-IV VCI, WMI, and FSIQ ($p < .05$). The NON-ADHD-REF group scored significantly lower than Control on WISC-IV PRI, WMI, and FSIQ ($p < .05$). There were no significant differences between the ADHD-C and ADHD-IN groups on any WISC-IV scale. There were no significant effects of gender or Gender \times Group.

Group Differences on TOF and WIAT-II

TOF. The top part of Table 3 shows means and standard deviations on the TOF scales for the four diagnostic groups based on examiners' ratings of children's behavior during the WIAT-II. The TOF results based on the WIAT-II were very similar to those based on the WISC-IV. The overall MANOVA for the five TOF syndromes showed a significant main effect of group ($F[15, 453] = 2.97; p < .001$; partial $\eta^2 = .08$). Subsequent ANOVAs revealed significant group effects ($p < .001$) only for Oppositional and Attention Problems, with medium to large ES (partial $\eta^2 = .14$ and $.15$, respectively). The 2×4 ANOVA for the ADHP scale total score revealed a significant main effect of group, with a large ES (partial $\eta^2 = .20$). The 2×4 MANOVA for the ADHP Inattention and Hyperactivity-Impulsivity subscales showed a significant main effect of group ($F[6, 334] = 6.60; p < .001$; partial $\eta^2 = .11$), and subsequent ANOVAs revealed significant group effects ($p < .001$) for both subscales, with medium to large ES (partial $\eta^2 = .15$ and $.18$). The 2×4 MANOVA for Internalizing and Externalizing showed a significant main effect of group ($F[6, 334] = 6.62; p < .001$; partial $\eta^2 = .11$), and subsequent ANOVAs revealed significant group effects ($p < .001$) only for Externalizing, with a large ES (partial $\eta^2 = .18$). Finally, the 2×4 ANOVA for Total Problems revealed a significant main effect of group, with a medium ES (partial $\eta^2 = .12$).

Tukey HSD tests showed that the ADHD-C group scored significantly higher than ADHD-IN, NON-ADHD-REF, and Control on TOF Oppositional, Attention Problems, the ADHP scale total score, ADHP Inatten-

Table 3
Group Differences on TOF and WIAT-II Scales

| Dependent Measures | ADHD-C (n = 73) | ADHD-IN (n = 25) | NON-ADHD-REF (n = 52) | Control (n = 26) | F | p | η^2 |
|----------------------------|------------------------------|------------------------------|------------------------------|-----------------------------|-------------------|-----------|----------|
| TOF scales | | | | | | | |
| Withdrawn/Depressed | 0.97 (3.59) | 2.00 (3.22) | 1.98 (4.06) | 2.19 (3.46) | <i>ns</i> | <i>ns</i> | .02 |
| Language/Thought Problems | 1.25 (1.92) | 1.64 (2.34) | 1.33 (2.19) | 0.62 (1.10) | <i>ns</i> | <i>ns</i> | .02 |
| Anxious | 3.81 (3.65) | 3.60 (2.78) | 4.37 (4.46) | 3.04 (3.22) | <i>ns</i> | <i>ns</i> | .01 |
| Oppositional | 5.60 (6.23) _a | 2.00 (2.06) _b | 1.94 (4.82) _b | 0.54 (1.10) _b | F(3, 168) = 9.02 | < .001 | .14 |
| Attention Problems | 12.27 (8.65) _a | 5.56 (5.47) _b | 7.23 (6.28) _b | 3.73 (4.96) _b | F(3, 168) = 10.23 | < .001 | .15 |
| ADHP total score | 15.19 (10.69) _a | 7.40 (6.49) _b | 7.62 (7.20) _b | 3.77 (4.96) _b | F(3, 168) = 13.84 | < .001 | .20 |
| ADHP Inattention | 5.89 (5.43) _a | 2.88 (3.88) _b | 3.13 (3.76) _b | 1.08 (1.60) _b | F(3, 168) = 9.66 | < .001 | .15 |
| ADHP Hyperactivity- | | | | | | | |
| Impulsivity | 9.30 (6.40) _a | 4.52 (3.50) _b | 4.48 (4.22) _b | 2.69 (4.01) _b | F(3, 168) = 12.48 | < .001 | .18 |
| Internalizing | 2.22 (4.25) | 3.64 (4.42) | 3.31 (4.61) | 2.81 (3.73) | <i>ns</i> | <i>ns</i> | .01 |
| Externalizing | 17.88 (13.58) _a | 7.56 (6.65) _b | 9.17 (9.21) _b | 4.27 (5.33) _b | F(3, 168) = 12.40 | < .001 | .18 |
| Total Problems | 35.53 (23.66) _a | 22.48 (12.89) _b | 25.06 (20.41) _{a,b} | 14.23 (10.68) _b | F(3, 168) = 7.31 | < .001 | .12 |
| WIAT-II scales | (n = 65) | (n = 23) | (n = 49) | (n = 24) | | | |
| RC | 96.32 (14.76) _a | 97.91 (12.53) _{a,b} | 106.71 (15.82) _b | 116.58 (11.16) _c | F(3, 153) = 11.73 | < .001 | .19 |
| MC | 97.40 (13.20) _{a,b} | 95.48 (8.29) _a | 106.20 (16.92) _b | 116.42 (15.24) _c | F(3, 153) = 14.34 | < .001 | .22 |
| WLC | 98.51 (17.92) _a | 97.61 (14.39) _a | 107.29 (15.51) _a | 120.96 (15.97) _b | F(3, 153) = 11.39 | < .001 | .18 |

Note. ADHD-C = attention deficit hyperactivity disorder—combined type; ADHD-IN = ADHD—inattentive type; NON-ADHD-REF = non-ADHD clinically referred; Control = controls; ADHP = Attention Deficit/Hyperactivity Problems; TOF = Test Observation Form; WIAT-II = Wechsler Individual Achievement Tests—Second Edition; RC = Reading Composite; MC = Mathematics Composite; WLC = Written Language Composite. For TOF, total N = 176; for WIAT-II, total N = 161 with complete data. Mean (SD). TOF scores are raw scores; WIAT-II scores are standard scores. Means in the same row with the same subscripts do not differ significantly at $p < .05$ by Tukey HSD comparisons of homogeneous subsets.

tion, ADHP Hyperactivity-Impulsivity, and Externalizing (all $p < .05$). On TOF Total Problems, ADHD-C scored significantly higher than ADHD-IN and Control ($p < .05$), but not higher than NON-ADHD-REF. There were no significant differences between the ADHD-IN versus NON-ADHD-REF and Control groups on any TOF scale. There were no significant effects of gender or Gender \times Group.

WIAT-II. The bottom part of Table 3 shows means and standard deviations for the four diagnostic groups on WIAT-II composite scores. The overall MANOVA for the three WIAT-II composite scores showed a significant main effect of group ($F[9, 368] = 5.50$; $p < .001$; partial $\eta^2 = .10$). Subsequent ANOVAs revealed significant group effects ($p < .001$) for all WIAT-II composite scores, with large ES (partial $\eta^2 = .18$ to $.22$).

Tukey HSD tests showed that both the ADHD-C and ADHD-IN groups scored significantly lower than Control on all three WIAT-II composite scores ($p < .05$). ADHD-C scored lower than NON-ADHD-REF on WIAT-II RC ($p < .05$), but not MC or WLC. The ADHD-IN group scored significantly lower than NON-ADHD-REF on WIAT-II MC ($p < .05$), but not RC or WLC. The NON-ADHD-REF group scored significantly lower than Control on all three WIAT-II composites ($p < .05$). There were no significant differences between the ADHD-C and ADHD-IN groups on any WIAT-II composite. There were no significant effects of gender or Gender \times Group.

FSIQ as a covariate. Because there were significant group differences on WISC-IV FSIQ, we reran the analyses of TOF scores reported in Tables 2 and 3, using 2×4 multivariate analyses of covariance (MANCOVAs) and analyses of covariance (ANCOVAs), treating FSIQ as a covariate. Results produced slight reductions in ES for group main effects, but no changes in the patterns of main effects or interactions. The FSIQ covariate was significant in analyses of only one TOF scale (ADHP Inattention) based on examiners' ratings during the

WISC-IV and of three TOF scales (Attention Problems, ADHP Inattention, and the ADHP scale total score) based on examiners' ratings during the WIAT-II, but all ES values for the FSIQ covariate were small (partial $\eta^2 = .03$ to $.05$). Subsequent ANCOVAs showed no changes in group differences on four TOF scales. ADHD-C continued to score significantly higher than ADHD-IN, NON-ADHD-REF, and Control on Attention Problems, the ADHP scale total score, ADHP Inattention, and Externalizing for TOF ratings based on both the WISC-IV and WIAT-II. There were changes in group differences on ADHP Hyperactivity-Impulsivity and Total Problems. For TOF ratings based on the WISC-IV, ADHD-C no longer scored significantly higher than ADHD-IN on Oppositional and ADHP Hyperactivity-Impulsivity, but ADHD-C now scored significantly higher than NON-ADHD-REF as well as Control on Total Problems. For TOF ratings based on the WIAT-II, ADHD-C no longer scored significantly higher than ADHD-IN on Total Problems.

Learning disability (LD) as a covariate. We were also interested to learn whether the presence or absence of a learning disability would change the direction of group effects on the TOF scales and WISC-IV and WIAT-II scores. Children were identified as having or not having LD according to a proxy variable derived from the discrepancy between WISC-IV FSIQ and predicted achievement scores for WIAT-II RC, MC, and WLC, following the predicted achievement method (versus simple discrepancy) recommended in the WIAT-II examiner's manual (Wechsler, 2002, pp. 155–158). Children were identified as having or not having LD (coded 1, 0) if the difference between predicted achievement and actual achievement was higher than the difference score at $p < .01$ on any one of the three WIAT-II composites. The discrepancy method identified 28.8% ($n = 51$) of the total sample of 177 as having LD: 36.5% ($n = 27$) for ADHD-C; 36% ($n = 9$) for ADHD-IN; 26.9% ($n = 14$) for NON-ADHD-REF; and 3.8% ($n = 1$) for Controls. There were 10 cases (5.6%) with missing data for LD, reducing the

total sample to 167 for analyses with the LD covariate.

To test the effects of LD on ratings of test session behavior, we reran analyses of TOF scores reported in Tables 2 and 3, using 2×4 MANCOVAs and ANCOVAs treating LD as a covariate. Results from all analyses of TOF scales again produced only slight reductions in ES for significant group main effects, and no changes in the patterns of main effects or interactions. The LD covariate was significant in analyses of three TOF scales (the ADHP scale total score, ADHP Inattention, and Total Problems) based on examiners' ratings during the WISC-IV and of five TOF scales (Oppositional, ADHP Inattention, Internalizing, Externalizing, and Total Problems) based on examiners' ratings during the WIAT-II, but all ES were small (partial $\eta^2 = .03$ to $.05$). Subsequent ANCOVAs showed no changes in group differences on five TOF scales. For TOF ratings based on both the WISC-IV and WIAT-II, ADHD-C continued to score significantly higher than ADHD-IN, NON-ADHD-REF, and Control on Attention Problems, the ADHP scale total score, ADHP Inattention, ADHP Hyperactivity-Impulsivity, and Externalizing. There were only minor changes in group differences on two TOF scales. For TOF ratings based on the WISC-IV, ADHD-C no longer scored significantly higher than ADHD-IN on the TOF Oppositional syndrome. For TOF ratings based on both the WISC-IV and WIAT-II, ADHD-C now scored significantly higher than NON-ADHD-REF on Total Problems. For TOF ratings based on the WIAT-II, ADHD-C no longer scored significantly higher than ADHD-IN on Total Problems (because of a large increase in standard error).

When we reran analyses of WISC-IV scores, the LD covariate was significant only for VCI with a small ES ($\eta^2 = .03$), but there were no changes in the pattern of group effects on WISC-IV VCI. There were no changes in the patterns of group effects for ADHD-C versus other groups on any WISC-IV scales. ADHD-C continued to score significantly lower than Control on all WISC-IV scales, and lower than NON-ADHD-REF on WISC-IV

VCI, WMI, and FSIQ. There continued to be no significant differences between ADHD-C and ADHD-IN on any WISC-IV scale. There were slight changes in group differences involving ADHD-IN. ADHD-IN no longer scored significantly lower than Control on WISC-IV PRI and PSI and no longer scored significantly lower than NON-ADHD-REF on WISC-IV WMI. ADHD-IN continued to score significantly lower than Control on WISC-IV WMI.

When we reran analyses of WIAT-II scores, the LD covariate was significant on all three composite scores with medium to large ES (partial $\eta^2 = .13$ to $.20$). ADHD-C and ADHD-IN continued to score significantly lower than Control on all three WIAT-II composite scores. ADHD-C continued to score significantly lower than NON-ADHD-REF on WIAT-II RC. On WIAT-II MC, ADHD-C now scored significantly lower than NON-ADHD-REF, whereas ADHD-IN no longer scored significantly lower than NON-ADHD-REF. There continued to be no significant differences between ADHD-C and ADHD-IN on any WIAT-II composite score. There were no significant differences between NON-ADHD-REF and Control on any WIAT-II composites.

SES as a covariate. Because there were significant group differences and site differences on SES, we reran our analyses of TOF scores in Tables 2 and 3, using 2×4 MANCOVAs and ANCOVAs, treating SES as a covariate. Results showed no significant effects of SES as a covariate for any TOF scales based on ratings for both the WISC-IV and WIAT-II. There was only one change in group effects, with ADHD-C no longer scoring significantly higher than ADHD-IN on TOF Total Problems for ratings based on the WIAT-II.

Discriminant Analyses

ADHD-C versus NON-ADHD-REF. Table 4 shows the combinations of significant predictors producing the best cross-validated classification rates for ADHD-C versus NON-ADHD-REF. The first column shows the sig-

Table 4
Cross-Validated Percentages of Cases Correctly Classified as ADHD-C
Versus NON-ADHD-REF

| Significant Predictors | Standardized Canonical Coefficients | ADHD-C (<i>n</i> = 74/65) ^a | NON-ADHD-REF (<i>n</i> = 52/49) ^a | Overall Correct Classification ^b |
|---|---|--|--|---|
| a. TOF Attention Problems for WISC-IV | 1.095 | 71.6% | 78.8% | 74.6% |
| TOF Anxious for WISC-IV | -0.631 | | | |
| b. TOF Attention Problems for WISC-IV | -0.867 | 64.9% | 73.1% | 68.3% |
| WISC-IV VCI | 0.482 | | | |
| TOF Anxious for WISC-IV | 0.472 | | | |
| c. TOF ADHP Hyperactivity- Impulsivity for WISC-IV | 0.776 | 64.9% | 75.0% | 69.0% |
| WISC-IV VCI | -0.623 | | | |
| d. TOF ADHP total score for WISC-IV | 0.750 | 63.5% | 76.9% | 69.0% |
| WISC-IV VCI | -0.598 | | | |
| e. WIAT-II RC | 0.685 | 73.5% | 72.0% | 72.9% |
| TOF Oppositional for WIAT-II | -0.671 | | | |
| f. TOF ADHP Hyperactivity- Impulsivity for WIAT-II | 0.772 | 72.1% | 72.0% | 72.0% |
| WIAT-II RC | -0.568 | | | |
| g. TOF ADHP total score for WIAT-II | 0.734 | 66.2% | 70.0% | 67.8% |
| WIAT-II RC | -0.563 | | | |

Note. All discriminant functions were significant at $p < .001$.¹ ADHD-C = attention deficit hyperactivity disorder—combined type; NON-ADHD-REF = non-ADHD clinically referred; ADHP = Attention Deficit/Hyperactivity Problems; TOF = Test Observation Form; WIAT-II = Wechsler Individual Achievement Tests—Second Edition; WISC-IV = Wechsler Intelligence Scale for Children—Fourth Edition; VCI = Verbal Comprehension Index; RC = Reading Composite.

^a *n* for TOF and WISC-IV/*n* for TOF and WIAT-II.

^b Percentage of cross-validated grouped cases correctly classified.

nificant predictors in the discriminant function.² When the five TOF syndromes were entered as candidate predictors (a), Attention Problems and Anxious emerged as significant predictors, correctly classifying 71.6% as ADHD-C and 78.8% as NON-ADHD-REF, with an overall correct classification of 74.6%. Standardized canonical coefficients indicated that Attention Problems was a larger contributor to group classification than Anxious and that the contribution of these two scales was reversed. When the four WISC-IV Index scores were added as candidate predictors

along with TOF scales (b to d), WISC-IV VCI emerged as an additional significant predictor along with TOF Attention Problems and Anxious, ADHP Hyperactivity-Impulsivity, or the ADHP scale total score, but classification rates were somewhat reduced. When the three WIAT-II composite scores were added as candidate predictors along with TOF scales (e to g), WIAT-II RC emerged as a significant predictor along with one TOF scale. TOF Oppositional, and WIAT-II RC (e) and TOF ADHP Hyperactivity-Impulsivity and WIAT-II RC (f) produced classification

Table 5
Cross-Validated Percentages of Cases Correctly Classified as ADHD-C
Versus Control

| Significant Predictors | Standardized Canonical Coefficients | ADHD-C (<i>n</i> = 74/65) ^a | Control (<i>n</i> = 26/24) ^a | Overall Correct Classification ^b |
|--|---|--|---|---|
| a. WISC-IV WMI | 0.526 | 77.0% | 84.6% | 79.0% |
| TOF Attention Problems for WISC-IV | -0.490 | | | |
| WISC-IV VCI | 0.395 | | | |
| WISC-IV PSI | 0.340 | | | |
| b. TOF ADHP Hyperactivity-Impulsivity for WISC-IV | -0.519 | 78.4% | 84.6% | 80.0% |
| WISC-IV WMI | 0.491 | | | |
| WISC-IV VCI | 0.446 | | | |
| WISC-IV PSI | 0.338 | | | |
| c. WISC-IV WMI | 0.504 | 79.7% | 80.8% | 80.0% |
| TOF ADHP total score for WISC-IV | -0.487 | | | |
| WISC-IV VCI | 0.417 | | | |
| WISC-IV PSI | 0.332 | | | |
| d. WIAT-II RC | 0.485 | 80.9% | 91.7% | 83.7% |
| TOF Attention Problems for WIAT-II | -0.456 | | | |
| WIAT-II MC | 0.443 | | | |
| e. TOF ADHP Hyperactivity-Impulsivity for WIAT-II | -0.541 | 82.4% | 91.7% | 84.8% |
| WIAT-II RC | 0.469 | | | |
| WIAT-II MC | 0.435 | | | |
| f. TOF ADHP total score for WIAT-II | -0.513 | 79.4% | 95.8% | 83.7% |
| WIAT-II RC | 0.454 | | | |
| WIAT-II MC | 0.421 | | | |

Note. All discriminant functions were significant at $p < .001$. ADHD-C = attention deficit hyperactivity disorder—combined type; Control = controls; ADHP = Attention Deficit/Hyperactivity Problems; TOF = Test Observation Form; WIAT-II = Wechsler Individual Achievement Tests-Second Edition; WISC-IV = Wechsler Intelligence Scale for Children—Fourth Edition; WMI = Working Memory Index; VCI = Verbal Comprehension Index; PSI = Processing Speed Index; RC = Reading Composite; MC = Mathematics Composite.

^a *n* for TOF and WISC-IV/*n* for TOF and WIAT-II.

^b Percentage of cross-validated grouped cases correctly classified.

rates over 70% for both ADHD-C and NON-ADHD REF.

ADHD-C versus Control. Table 5 shows the combinations of significant predictors producing the best classification rates for ADHD-C versus Control. Combinations of TOF scales and WISC-IV Index scores or TOF scales and WIAT-II composite scores correctly classified 77% to 82.4% of ADHD-C and 80.8% to 95.8% of Control cases, with overall correct classification rates of 79%

to 84.8%. TOF Attention Problems, ADHP Hyperactivity-Impulsivity, and the ADHP scale total score emerged as significant predictors, along with either WISC-IV VCI, WMI, and PSI (a to c), or WIAT-II RC and MC (d to f).

ADHD-C versus ADHD-IN. In discriminant analyses of ADHD-C versus ADHD-IN, only TOF scales emerged as significant predictors of group membership, correctly classifying 50% to 68.5% of ADHD-C

and 68% to 84% of ADHD-IN, with overall correct classification of 56.6% to 70.4%. No WISC-IV Index scores or WIAT-II composite scores emerged as significant predictors. The best combination of predictors was TOF Attention Problems and Language/Thought Problems in reverse direction, correctly classifying 68.5% of ADHD-C and 76% of ADHD-IN, with overall correct classification of 70.4%. Other discriminant analyses also showed TOF Attention Problems, the ADHP scale total score, ADHP Inattention, and ADHP Hyperactivity-Impulsivity as significant predictors.

ADHD-IN versus NON-ADHD-REF.

Discriminant analyses showed only WISC-IV VCI and WIAT-II MC as significant predictors of ADHD-IN versus NON-ADHD-REF. WISC-IV VCI correctly classified 80% of ADHD-IN and 63.5% of NON-ADHD-REF, with overall correct classification of 68.8%. WIAT-II MC correctly classified 76% of ADHD-IN and 57.7% of NON-ADHD-REF, with overall correct classification of 63.6%.

ADHD-IN versus Control. Discriminant analyses showed TOF Oppositional plus WISC-IV VCI and WMI as the best predictors of ADHD-IN versus Control, correctly classifying 88% of ADHD-IN and 84.6% of Control, with overall correct classification of 86.3%. In other analyses, WISC-IV VCI and WMI or WIAT-II MC and WLC were significant predictors without any TOF scale. WISC-IV VCI and WMI correctly classified 88% of ADHD-IN and 80.8% of Control, with overall correct classification of 84.3%. WIAT-II MC and WLC correctly classified 87% of ADHD-IN and 76% of Control, with overall correct classification of 81.3%.

Discussion

In the absence of valid laboratory or neuropsychological measures for diagnosing ADHD, direct observation of children's behavior can be an important adjunct to parent and teacher reports for assessing the core symptoms of ADHD. Test sessions offer a venue for direct observations that is available to both school psychologists and clinic-based

practitioners. Moreover, it is advisable to administer cognitive and achievement tests to children who may have ADHD in order to determine whether they have cognitive impairments, comorbid LD, or academic deficits that warrant academic interventions along with behavioral interventions (DuPaul & Stoner, 2003; DuPaul, 2007). In the present study, we used a standardized form, the TOF, to obtain systematic ratings of test session behavior for children with *DSM-IV* diagnoses of the ADHD—Combined and ADHD—Predominantly Inattentive subtypes versus other clinically referred children without ADHD and nonreferred control children.

ADHD-C Versus NON-ADHD-REF and Control

Consistent with our hypotheses, our results showed that the ADHD-C group scored significantly higher than NON-ADHD-REF and Control on the four TOF scales measuring ADHD-consistent problems of inattention and hyperactivity-impulsivity: Attention Problems, the ADHP scale total score, and the ADHP Inattention and Hyperactivity-Impulsivity subscales. These results were consistent across test session observations during both the WISC-IV and WIAT-II, with large ES (15% to 20% of variance), according to Cohen's (1988) criteria. The significant group differences also persisted, with little change in ES, even when FSIQ, LD, or SES were entered as covariates in subsequent analyses. These findings are especially notable given that test examiners were kept blinded to all referral and diagnostic information about the children, which has not been the case in previous studies (e.g., Solanto et al., 2007).

The ADHD-C group also scored significantly higher than NON-ADHD-REF and Control on the TOF Oppositional syndrome and Externalizing, with medium to large ES (13% to 18% of variance). These findings are consistent with the higher rates of parent-reported comorbid *DSM-IV* diagnoses of conduct disorder and oppositional defiant disorder for ADHD-C (16.2% and 52.7%, respectively) than for the NON-ADHD-REF (0% and 26.9%,

respectively) and Control (3.8% for each diagnosis). Although analyzing effects of comorbid diagnoses on test behavior was beyond the scope of our study, researchers in the NIMH Multisite Multimodal Treatment Study of ADHD have documented significant differences in functional impairment and response to treatment for children with ADHD who also have comorbid disruptive disorders and/or anxiety disorders (Jensen et al., 2001).

In addition to showing significantly more behavior problems during testing, the ADHD-C group scored significantly lower than Control on all four WISC-IV Index scores and FSIQ and all three WIAT-II composite scores. ADHD-C scored significantly lower than NON-ADHD-REF on WISC-IV VCI, WMI, and FSIQ and WIAT-II RC. These findings are consistent with many other studies showing significantly lower cognitive functioning and lower achievement for children with ADHD compared to other clinically referred children and typically developing children (Faroane, Biederman, Weber, & Russell, 1998; Frazier, Demaree, & Youngstrom, 2004; Mayes & Calhoun, 2007).

Discriminant analyses also revealed that test session observations, combined with cognitive and achievement test scores, significantly differentiated ADHD-C from clinically referred children without ADHD and typically developing controls. Specifically, TOF Attention Problems, the ADHP scale total score, or ADHP Hyperactivity-Impulsivity, combined with WISC-IV WMI, VCI and PSI, or WIAT-II RC and MC, were significant predictors of ADHD-C versus Control, correctly classifying $\geq 77\%$ of ADHD-C and $>80\%$ of Control. The ADHP scale total score or ADHP Hyperactivity-Impulsivity, along with WISC-IV VCI or WIAT-II RC, were also significant predictors of ADHD-C versus NON-ADHD-REF, correctly classifying $>63\%$ of ADHD-C and $\geq 70\%$ of NON-ADHD-REF. Interestingly, TOF Attention Problems and Anxious from observations during the WISC-IV produced the best classification rates for ADHD-C versus NON-ADHD-REF (71.6% and 78.8%, respectively), suggesting that relatively higher attention problems, combined with relatively lower

anxiety during the WISC-IV, might be important clinical markers for differentiating children with ADHD-C from other clinically referred children without ADHD. TOF Oppositional and WIAT-II RC were also significant predictors for ADHD-C versus NON-ADHD-REF, suggesting that oppositional behavior and low reading achievement may also be important clinical markers for differentiating children with ADHD from other clinically referred children.

The higher classification rates for ADHD-C versus Control than for ADHD-C versus NON-ADHD-REF are to be expected, because differentiating children with ADHD-C from other clinically referred children without ADHD is a more challenging clinical task than differentiating children with ADHD from typically developing children. With this in mind, our findings offer encouraging support for test session observations as external validators of parent and teacher reports for differential diagnosis of ADHD. Our findings also support clinical wisdom that places great value on observations of test session behavior along with interpretations of test scores in the assessment process (Sattler, 2008). Moreover, the TOF provides a standardized method for obtaining systematic observations during test sessions and quantifying the results of those observations.

ADHD-C Versus ADHD- IN Subtypes

An even more challenging clinical task is differentiating between the *DSM-IV* ADHD-C and ADHD-IN subtypes. Consistent with our hypotheses, we found that the ADHD-C group scored significantly higher than ADHD-IN on TOF Attention Problems, the ADHP scale total score, and the ADHP Hyperactivity-Impulsivity subscale, for ratings based on both the WISC-IV and WIAT-II. Surprisingly, the ADHD-C group also scored significantly higher than ADHD-IN on the ADHP Inattention subscale for ratings based on both tests, which was contrary to our hypotheses. This finding suggests that, at least in our sample of 6- to 11-year-old children, the ADHD-C group may represent a more severe version of ADHD than does the ADHD-IN group. Moreover, the significant differences between ADHD-C and

ADHD-IN on TOF Attention Problems, the ADHP scale total score, and ADHP Inattention, based on either the WISC-IV or WIAT-II, continued to appear even when FSIQ and LD were entered into analyses as covariates.

At the same time, we found no significant differences between ADHD-C and ADHD-IN on the four WISC-IV Index scores and FSIQ or the three WIAT-II composite scores. Discriminant analyses revealed that *test session observations alone* were the best predictors of ADHD-C versus ADHD-IN, with TOF Attention Problems, Language/Thought Problems, the ADHP scale total score, ADHP Inattention, and ADHP Hyperactivity-Impulsivity emerging as significant predictors in different sets of analyses. Interestingly, TOF Attention Problems and Language/Thought Problems based on the WIAT-II, in reverse direction, were the best predictors, correctly classifying 68.5% of ADHD-C and 76% of ADHD-IN. These findings, along with significant mean group differences, support the discriminative validity of test session observations for differentiating between the two ADHD subtypes.

The null findings for ADHD-C versus ADHD-IN on the WISC-IV or WIAT-II are consistent with other research studies showing few, if any, differences between the ADHD subtypes in cognitive functioning, academic achievement, or laboratory and neuropsychological tests (Chhabildas et al., 2001; Nigg et al., 2002; Solanto et al., 2007). It was notable, in particular, that Solanto et al. (2007) concluded that, after controlling for FSIQ, the *DSM-IV* ADHD-C, and ADHD-IN subtypes were best differentiated by parent and teacher rating scales and observational measures of children's off-task and impulsive behavior during cognitive tasks and a CPT. However, test examiners in the Solanto et al. study were not blinded to the children's diagnostic group status as were examiners in the present study.

ADHD-IN Versus NON-ADHD-REF and Control

In contrast to the previously mentioned findings, test session observations were not strong

predictors of ADHD-IN versus other clinically referred children without ADHD or typically developing controls. In discriminant analyses, the TOF Oppositional syndrome emerged as a significant predictor of ADHD-IN versus Control, but it was a weaker predictor than WISC-IV VCI and WMI. In other discriminant analyses, only WISC-IV VCI and WMI or WIAT-II MC and WLC emerged as significant predictors of ADHD-IN versus Control, and only WISC-IV VCI or WIAT-II MC emerged as predictors of ADHD-IN versus NON-ADHD-REF.

In a similar fashion, ADHD-IN differed from Control and NON-ADHD-REF only in lower mean scores on the WISC-IV and WIAT-II, but not in higher mean scores on the TOF. However, even these results must be viewed with some caution because certain group differences involving ADHD-IN on WISC-IV WMI, PRI, and PSI, and WIAT-II MC disappeared when LD was entered as a covariate, which was not the case for group differences involving ADHD-C. These findings suggest that even though the two ADHD subtypes had the same proportion of cases with comorbid LD (36%), LD had a greater effect on WISC-IV and WIAT-II performance for ADHD-IN than for ADHD-C. This is consistent with Mayes and Calhoun's (2007) finding that WISC-IV WMI and PSI were strong predictors of LD in their sample of children with ADHD.

Limitations

There are several limitations to our study. First, sample sizes were unequal for the four groups and *n* values were relatively small for ADHD-IN and Control. This could have reduced power for finding differences between ADHD-IN and other groups. However, we did take unequal *n* values into account by examining homogeneous subsets based on the harmonic mean sample size in post hoc Tukey HSD pairwise comparisons. Second, our sample was limited to 6- to 11-year-old children, so the results may not generalize to adolescents. Third, test examiners may have developed hypotheses about the children during testing that could have affected their ratings on the TOF. To minimize such rater effects, exam-

iners were kept blinded to all clinical information about the children; they were instructed to complete the TOF before scoring any tests; and they were provided behavioral descriptors as guidelines for scoring the TOF items.

Conclusions

In summary, medium to large group effects and good classification rates in this study provide strong evidence of the discriminative validity of test session observations, as measured on the TOF, for differentiating children with ADHD—Combined subtype from other clinically referred children without ADHD and typically developing children. The best observational measures of ADHD—Combined were TOF scales assessing attention problems and hyperactivity-impulsivity consistent with parent and teacher-reported ADHD symptoms. The discriminative power of the TOF scales for ADHD—Combined versus the other two groups without ADHD was retained even after adding FSIQ, LD, or SES as covariates in subsequent analyses and adding WISC-IV and WIAT-II scores as candidate predictors in discriminant analyses. Consistent with other research, children with ADHD—Combined showed significantly lower cognitive ability on the WISC-IV and lower achievement on the WIAT-II compared to typically developing children. They also showed lower cognitive ability and lower reading achievement than other clinically referred children without ADHD.

The same TOF scales showed good discriminative power for differentiating between the ADHD—Combined and ADHD—Predominately Inattentive subtypes, corroborating parent and teacher reports, although classification rates were lower. It was notable that only test session observations, and not WISC-IV or WIAT-II scores, differentiated between the ADHD subtypes in discriminant analyses. By contrast, test session observations did not corroborate parent and teacher reports in differentiating children with ADHD—Predominately Inattentive subtype from clinically referred children without ADHD or typically developing children. Instead, WISC-IV and WIAT-II test scores were

better discriminators of ADHD—Predominately Inattentive versus the latter two groups. Comorbidity of LD with ADHD—Predominately Inattentive was also an important factor to consider in differentiating ADHD—Predominately Inattentive from clinically referred children without ADHD and typically developing children.

Footnotes

¹To adjust for unequal sample sizes, the Tukey HSD test on homogenous subsets uses the harmonic mean sample size and $\alpha = .05$ to test group differences. However, multiple comparisons based on MANOVAs and univariate ANOVAs in our analyses showed most significant group differences at $p < .01$ or lower.

²All classification rates were obtained with prior probabilities equal for all groups.

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, and Families.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- 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.
- Barkley, R. A., DuPaul, G. J., & McMurray, M. B. (1990). A comprehensive evaluation of attention deficit disorder with and without hyperactivity. *Journal of Consulting and Clinical Psychology, 58*, 775–789.
- Brock, S. E., & Clinton, A. (2007). Diagnosis of attention-deficit/hyperactivity disorder (AD/HD) in childhood: A review of the literature. *The California School Psychologist, 12*, 73–91.
- Chhabildas, N., Pennington, B. F., & Willcutt, E. G. (2001). A comparison of the neuropsychological profiles of the DSM-IV subtypes of ADHD. *Journal of Abnormal Child Psychology, 29*, 529–540.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Demaray, M. K., Schaefer, K., & Delong, L. K. (2003). Attention deficit/hyperactivity disorder (ADHD): A national survey of training and assessment practices in the schools. *Psychology in the Schools, 40*, 583–597.
- DuPaul, G. J. (2007). School-based interventions for students with attention deficit hyperactivity disorder: Cur-

- rent status and future directions. *School Psychology Review*, 36, 183–194.
- 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., & Stoner, G. (2003). *ADHD in the schools* (2nd ed.) New York: Guilford Press.
- Faraone, S., Biederman, J., Webber, W., & Russell, R. (1998). Psychiatric, neuropsychological, and psychosocial features of *DSM-IV* subtypes of attention-deficit/hyperactivity disorder: Results from a clinically referred sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 185–193.
- Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention deficit/hyperactivity disorder. *Neuropsychology*, 18, 543–555.
- Gittleman, R. G., & Klein, D. (1985). Hillside Behavior Rating Scale. *Psychopharmacology Bulletin*, 21, 898–899.
- Glutting, J. J., & Oakland, T. (1993). *Manual for the Guide to the Assessment of Test Behavior*. San Antonio, TX: Psychological Corporation.
- Glutting, J. J., Robins, P. M., & de Lancy, E. (1997). Discriminant validity of test observations for children with attention deficit hyperactivity disorder. *Journal of School Psychology*, 35, 391–401.
- Glutting, J. J., Youngstrom, E. A., Oakland, T., & Watkins, M. W. (1996). Situational specificity and generality of test behaviors for samples of normal and referred children. *School Psychology Review*, 25, 94–107.
- Gordon, M., DiNiro, D., Mettelman, B. B., & Tallmadge, J. (1989). Observations of test behavior, quantitative scores, and teacher ratings. *Journal of Psychoeducational Assessment*, 7, 141–147.
- Hollingshead, A. B. (1975). *Four factor index of social status*. Unpublished paper. New Haven, CT: Yale University, Department of Sociology.
- Jensen, P. T., Hinshaw, S. P., Kraemer, H. C., Lenora, N., Newcorn, J. H., Abikoff, H. B., et al. (2001). ADHD comorbidity findings from the MTA Study: Comparing comorbid subtypes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 147–158.
- Mash, E. J., & Terdal, L. G. (Eds.). (2000). *Assessment of childhood disorders* (3rd ed.). New York: Guilford Press.
- Mayes, S. D., & Calhoun, S. L. (2007). Wechsler Intelligence Scale for Children—Third Edition and —Fourth Edition predictors of academic achievement in children with attention deficit/hyperactivity disorder. *School Psychology Quarterly*, 22, 234–249.
- McConaughy, S. H. (2005a). Direct observational assessment during test sessions and child clinical interviews. *School Psychology Review*, 34, 490–506.
- McConaughy, S. H. (2005b). The Test Observation Form (TOF): A new observational tool for school psychologists. *Communiqué*, 33, 36–38.
- McConaughy, S. H., & Achenbach, T. M. (2004). *Manual for the Test Observation Form for Ages 2–18*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- National Institutes of Health. (2000). Consensus Development Conference statement: Diagnosis and treatment of attention-deficit/hyperactivity disorder (ADHD). *Journal of the American Academy of Child and Adolescent Psychiatry*, 39, 182–193.
- Nigg, J. T. (2006). *What causes ADHD?* New York: Guilford Press.
- Nigg, J. T., Blaskey, L., Huang-Pollack, C., & Rappley, M. D. (2002). Neuropsychological executive functions and *DSM-IV* ADHD subtypes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 59–66.
- Reid, R., Maag, J. W., & Vasa, S. F. (1993). Attention deficit hyperactivity disorder as a disability category: A critique. *Exceptional Children*, 60, 198–214.
- Sattler, J. M. (2008). *Assessment of children: Cognitive applications* (5th ed.). La Mesa, CA: Author.
- Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children, Version IV (NIMH DISC-IV): Description, differences from previous versions and reliability for some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39, 28–38.
- Shapiro, E. S., & Kratochwill, T. R. (Eds.). (2000). *Conducting school-based assessments of child and adolescent behavior*. New York: Guilford Press.
- Solanto, M. V., Gilbert, S. N., Raj, A., Zhu, J., Pope-Boyd, S., Stepak, B., et al. (2007). Neurocognitive functioning in AD/HD, predominantly inattentive and combined subtypes. *Journal of Abnormal Child Psychology*, 35, 729–744.
- Teicher, M. H., Ito, Y., Glod, C. A., & Barber, N. I. Objective measurement of hyperactivity and attentional problems in ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35, 334–342.
- Thorndike, R., Hagan, E., & Sattler, J. (1986). *The Stanford Binet Intelligence Scale* (4th ed.) Chicago: Riverside Press.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1992). *Wechsler Individual Achievement Test*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2002). *Wechsler Individual Achievement Tests—Second Edition*. San Antonio, TX: Psychological Corporation.
- Willcutt, E. G., Hartung, C. M., Lahey, B. B., Loney, J., & Pelham, W. E. (1999). Utility of behavior ratings by examiners during assessments of preschool children with attention-deficit/hyperactivity disorder. *Journal of Abnormal Child Psychology*, 27, 463–472.

Date Received: August 5, 2008

Date Accepted: December 29, 2008

Action Editor: Tanya Eckert ■

Stephanie H. McConaughy, PhD, is Research Professor of Psychiatry and Psychology at the University of Vermont. She is a Vermont-licensed practicing psychologist and Nationally Certified School Psychologist. Her research focuses on empirically based assessment of children's behavioral and emotional problems, clinical interviewing, and school-based prevention programs for behavioral disorders.

Masha Y. Ivanova is Research Assistant Professor of Psychiatry at the University of Vermont. She received her PhD at the University at Albany and completed her predoctoral clinical internship and postdoctoral training at the University of Vermont Center for Children, Youth, and Families. Her research focuses on the understanding of environmental factors as risk and protective factors for child psychopathology.

Kevin Antshel, PhD, is Assistant Professor of Psychiatry and Director of the Adult ADHD Treatment & Research Program at the State University of New York (SUNY)—Upstate Medical University. His research and clinical interests include ADHD, learning disabilities, and developmental neuropsychology.

Ricardo B. Eiraldi, PhD, is Assistant Professor of Clinical Psychology in the Department of Pediatrics at the University of Pennsylvania. His research focuses on the clinical presentation of ADHD in girls and ethnic minority populations; the application of help-seeking behavior models in the study of health disparities among ethnic minority children and families; and the development of strategies for addressing mental health services disparities in the inner city.