Neuropsychology
2004, Vol. 18, No. 3, 543–555

0894-4105/04/$12.00 DOI: 10.1037/0894-4105.18.3.543

Meta-Analysis of Intellectual and Neuropsychological Test Performance in Attention-Deficit/Hyperactivity Disorder

Thomas W. Frazier and Heath A. Demaree
Case Western Reserve University

Eric A. Youngstrom
Case Western Reserve University and University Hospitals of Cleveland

Cognitive measures are used frequently in the assessment and diagnosis of attention-deficit/hyperactivity disorder (ADHD). In this meta-analytic review, the authors sought to examine the magnitude of differences between ADHD and healthy participants on several commonly used intellectual and neuropsychological measures. Effect sizes for overall intellectual ability (Full Scale IQ; FSIQ) were significantly different between ADHD and healthy participants (weighted $d = .61$). Effect sizes for FSIQ were significantly smaller than those for spelling and arithmetic achievement tests and marginally significantly smaller than those for continuous performance tests but were comparable to effect sizes for all other measures. These findings indicate that overall cognitive ability is significantly lower among persons with ADHD and that FSIQ may show as large a difference between ADHD and control participants as most other measures.

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common psychiatric disorders in childhood, with an estimated incidence of 3%–5% (American Psychiatric Association, 1994). The problem is also of consequence for many adults who were either first diagnosed in adulthood or for whom childhood symptoms persist (Faraone et al., 2000; Woods, Lovejoy, & Ball, 2002). Improved understanding of the core cognitive, behavioral, and emotional consequences of the disorder has occurred over the last few decades. However, no diagnostic test exists to parse attention and concentration and self-regulation problems in ADHD from normal fluctuations in attention and self-regulation.

At present, clinical assessments of ADHD vary greatly from one clinician to another. Some individuals may be evaluated by clinicians using only a single information source, such as an unstructured interview with the patient or parent, whereas others may receive comprehensive assessments including cognitive and neuropsychological testing, standardized rating scales from multiple informants, behavioral observations, and structured diagnostic interviews. Comprehensive assessments are generally preferred over single-informant reports because they permit a thorough understanding of the patient's difficulties and offer the opportunity to rule out alternative explanations for the pattern of symptoms. Within the context of these assessments, cognitive and neuropsychological measures facilitate an unbiased evaluation of the core cognitive deficits that are observed in ADHD. However, there is little agreement in the literature regarding the differential utility of neurocognitive instruments (Rapport, Chung, Shore, Denney, & Isaacs, 2000). This is especially disturbing given the number of instruments available for assessing cognitive abilities. In the present meta-analytic review, we sought to address this gap in knowledge by determining and comparing the mean effect sizes for a wide range of intellectual and neuropsychological measures of ADHD.

Cognitive Assessment of ADHD

A typical starting point for neuropsychological assessments is the evaluation of an individual's overall level of cognitive ability. Researchers derive these estimates using tests of intellectual ability with an aggregate or Full Scale IQ (FSIQ) score serving as the omnibus estimate. In the context of an extensive evaluation of specific functions, overall estimates of ability provide a meaningful baseline for determining cognitive strengths and weaknesses. Many researchers have found significant decrements in overall cognitive ability in ADHD, with some studies finding as much as 20-point discrepancies between ADHD and healthy control groups (Abikoff, Courtney, Szeibel, & Koplewicz, 1996; Garcia-Sanchez, Estevez Gonzalez, Suarez Romero, & Junque, 1997). Other studies have found minimal, if any, differences between groups (Carlson, Mann, & Alexander, 2000; Carlson & Tamm, 2000; Kemner, Verbaten, Cuperus, Camfferman, & van Engeland, 1995; Ozonoff & Jensen, 1999). However, these nonsignificant intergroup differences may be due to a lack of statistical power and not to true equivalence. The first purpose of the present review was to use meta-analytic techniques to determine whether deficits in overall cognitive ability exist in ADHD and, if so, to characterize the magnitude of the deficit.

After determining the level of overall cognitive functioning, neuropsychological assessment serves the purpose of identifying the specific cognitive processes that are deficient in the individual or condition of interest. Although there is some disagreement about the exact nature of deficits, Barkley (1997a) proposed that

Thomas W. Frazier and Heath A. Demaree, Department of Psychology, Case Western Reserve University; Eric A. Youngstrom, Department of Psychology, Case Western Reserve University, and Department of Psychiatry, University Hospitals of Cleveland.

Additional materials are on the Web at http://dx.doi.org/10.1037/0894-4105.18.3.543.supp.

Correspondence concerning this article should be addressed to Thomas W. Frazier, Department of Psychology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7123. E-mail: twf3@po.cwru.edu. After September 1, 2004, correspondence should be addressed to Thomas W. Frazier, Department of Psychiatry and Psychology (P57), Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195. E-mail: fraziert@ccf.org

ADHD results in impairment involving behavioral inhibition processes (see also Quay, 1997). According to this theory, the core deficit of behavioral disinhibition contributes to disruptions of specific neuropsychological functions, including working memory, sustained attention, motor control, and affect regulation. Barkley's theory is supported by a growing number of studies that have found executive functioning deficits in hyperactive–impulsive or combined-subtype ADHD individuals. However, several studies have also reported deficits on non-executive functioning measures (Garcia-Sanchez et al., 1997; Rucklidge & Tannock, 2001), calling into question the specificity of neuropsychological deficits.

As a secondary objective, the present review attempted to meta-analytically specify the exact nature and magnitude of neuropsychological deficits in ADHD. Specifically, this review sought to determine the mean effect sizes for commonly used neuropsychological tests and compare these with the mean effect sizes for measures of estimated overall cognitive ability. Comparisons between IQ and neuropsychological tests, as well as among neuropsychological tests, will permit an evaluation of which measures show the largest group differences and thus, of which measures are most likely to provide incremental validity in the assessment of ADHD. Along with traditional neuropsychological measures, tests of academic achievement in spelling, reading, and arithmetic were included in the present review because these measures are frequently used to assess the presence of a learning disability, a common comorbid condition with ADHD (Dykman & Ackerman, 1991; Semrud-Clikeman et al., 1992). Inclusion of these achievement measures may also provide information regarding specific cognitive deficits.

It should be noted that Barkley's model is based on findings for hyperactive–impulsive or combined-type ADHD groups. Some research has suggested that a different pattern of neuropsychological deficits exists in inattentive-type ADHD groups. Specifically, Goodyear and Hynd (1992) and Barkley (1997b) posited that individuals with the inattentive subtype experience deficits in focused and selective attention and speed of information processing, whereas the other subtypes have deficits primarily in sustained attention, behavioral inhibition, and affect regulation. However, recent studies have found few, if any, differences between inattentive and combined-type groups (Chhabildas, Pennington, & Willcutt, 2001; Fischer, Barkley, Edelbrock, & Smallish, 1990; Klorman et al., 1999). For this reason, and because few studies in the present review included groups of inattentive participants ($n = 8$), analyses were expected mainly to reflect the intellectual and neuropsychological differentiation of the hyperactive–impulsive and combined subtypes from normals. Despite the small number of studies including groups of exclusively inattentive participants, preliminary analyses were performed examining subtype differences for FSIQ.

The Present Review

To address the objectives outlined above, we proposed the following hypotheses:

1. ADHD was expected to result in significant decrements in overall cognitive ability, as estimated by measures of FSIQ. Effect sizes for Verbal and Performance IQ subscales were expected to be roughly equivalent because both measures include tests that are sensitive to attention and working memory (Kaufman, 1994; Sattler, 2001). No specific predictions were made regarding differential magnitudes of FSIQ differences between ADHD subtypes.
2. All of the cognitive and neuropsychological measures examined in this review were expected to result in significant differences between ADHD and control participants, reflecting deficits in a wide range of cognitive functions.
3. Effect sizes for FSIQ were expected to be larger than those from neuropsychological measures of nonexecutive functions, but smaller than effect sizes from measures of executive functions. This was predicted because FSIQ is influenced by both executive and nonexecutive functions (for a review of executive and nonexecutive functions assessed by intellectual ability subtests, see Kaufman, 1990; Lezak, 1995).
4. Effect sizes from measures of executive functions were expected to be larger than effect sizes from measures of visual constructional ability, visual memory, verbal fluency, and receptive and expressive language functions. Specific predictions were not made regarding measures of academic functioning.

The latter two hypotheses are based on previous findings suggesting primarily executive-functioning deficits in individuals with ADHD.

Method

Literature Search

Inclusion and exclusion criteria. The literature concerning ADHD was searched with the PsycINFO and MEDLINE bibliographic databases. Only articles published during or after 1980 were reviewed, because these data coincide with the publication of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–III;* 3rd ed.; American Psychiatric Association, 1980) criteria for attention deficit disorder. PsycINFO and MEDLINE search terms included the following: *attention, attention deficit disorder, ADHD, sustained attention, vigilance, cognitive assessment,* and *neuropsychological; and* was included as an operator between search terms where relevant. An additional limitation that the articles had to be written in English was imposed. Dissertation abstracts were searched; however, none included sufficient information in the abstract to compute the relevant effect sizes. Additional articles were obtained through inspection of the reference lists of articles obtained in the above search. The search cutoff date was October 2002.

Only studies reporting data for overall cognitive ability, FSIQ, in sufficient detail for the calculation of effect sizes were included in the present review. This restriction was imposed because the primary hypotheses of the study involved examining the magnitude of FSIQ differences between ADHD participants and controls and comparing the effect sizes for FSIQ with the effect sizes for various neuropsychological measures. Studies without data for estimating FSIQ were not included because these could not be included in statistical tests of these hypotheses. Table 1 presents the authors and years of publication along with methodological characteristics and effect sizes for FSIQ for these studies. A total of 123 studies were identified. Several studies were excluded because of reported sample overlap with other included studies. In cases of sample overlap, studies were selected that provided the greatest amount of data (i.e., largest sample sizes or roughly equivalent sample sizes with more cognitive measures). Of the identified studies, 13 included multiple comparisons between control groups and ADHD subtypes. Although data sets with nested effect sizes are

Table 1
Demographics, Selected Methodological Characteristics, and Effect Sizes for FSIQ for All Comparisons

Study	*DSM*[a]	Control *n*	ADHD *n*	Control: % male	ADHD: % male	Control age (years)	ADHD age (years)	IQ test	IQ estimated[b]	IQ *d*
Abikoff et al., 1996	3	20	20	100	100	10.1	9.8	WISC–R	3	1.38
Assesmany et al., 2001	4 any	40	40	80	80	9.8	10.5	WISC–III	0	0.52
August & Garfinkel, 1990	3	43	70	100	100	10.7	12.3	PPVT	3	0.81
Barber et al., 1996	3	45	45	100	100	9.4	9.3	PPVT–R	3	0.21
Barkley et al., 1990	3 C	34	42	97	93	8.8	8.3	WISC–R	0	0.54
	3 I	34	48	97	90	8.8	9.0	WISC–R	0	0.61
Barkley et al., 1996	4 C	23	25	61	64	22.0	22.5	K–BIT	1	0.11
Barkley et al., 1997	4 unsp	26	12	—	—	10.5	11.2	—	—	0.48
Barry et al., 2001	4 C, I	15	15	100	100	10.4	10.5	K–BIT	1	0.74
Berquin et al., 1998	3	47	46	100	100	11.8	11.7	WISC–R	2	0.6
Boerger et al., 1999	3	16	21	100	100	8.4	8.8	WISC–R	0	1.07
Boucagnani et al., 1989	2 ADDH	28	28	86	86	—	—	WISC–R	1	0.45
Braaten & Rosen, 2000	4 C, H	19	24	100	100	7.8	8.3	PPVT–R	3	0.38
Breier et al., 2002	4 C	25	18	60	28	9.9	10.9	WASI	1	0.87
Bush et al., 1999	4 unsp	8	8	63	63	37.3	36.6	WAIS–R	1	0.55
Carlson & Tamm, 2000	4 C	22	22	68	73	9.6	9.7	WISC–R	1	0.07
Carlson et al., 2000	3	40	40	68	68	10.0	10.0	PPVT–R	3	0.11
Carte et al., 1996	3	31	51	100	100	9.0	9.6	WISC–R, WISC–III	2	0.36
Carter et al., 1995	3	20	20	75	75	10.7	10.5	WISC–R	0	0.56
Carter et al., 1995	3	19	19	74	79	10.6	10.6	WISC–R	0	0.63
Casey et al., 1997	3	26	26	100	100	9.8	9.7	WISC–R	1	0.48
Castellanos et al., 1996	3	55	57	100	100	12.0	11.7	WISC–R	1	0.67
Castellanos et al., 2001	4 C	50	50	0	0	10.0	9.7	WISC–R, WISC–III	1	0.49
Chee et al., 1989	2 ADDH	36	14	100	100	8.6	7.9	WISC–R	1	0.37
Chelune et al., 1986	2 ADDH	24	24	71	71	9.4	9.4	PPVT	3	0.29
Chhabildas et al., 2001	4 C	82	33	42	73	11.4	10.9	WISC–R, WAIS–R	0	1.38
	4 H	82	14	42	71	11.4	10.3	WISC–R, WAIS–R	0	0.11
	4 I	82	67	42	67	11.4	12.0	WISC–R, WAIS–R	0	1.37
Clark et al., 2000	4 any	26	35	77	86	14.0	14.2	WISC–III	1	1.33
Clarke et al., 2001	4 C, I	40	80	80	80	10.4	10.4	WISC–III	0	1.31
Clarke et al., 2001	4 C, I	80	264	73	73	10.4	10.3	WISC–III	0	0.75
Corbett & Glidden, 2000	4 unsp	37	37	51	76	9.5	10.1	WISC–III	0	1.01
Dane et al., 2000	4 C	22	22	64	82	9.1	9.1	—	—	1.11
	4 I	22	20	64	85	9.1	9.3	—	—	1.18
Dewey et al., 2001	3	112	53	73	87	11.3	12.4	WISC–III, WISC–R	1	0.53
Ernst et al., 1994	3	9	10	67	70	14.4	14.8	WISC–R	1	0.96
Ernst et al., 1999	3	10	10	70	80	14.8	13.8	WISC–R	1	0.92
Felton et al., 1987	2 ADDH	40	13	—	—	10.5	10.5	PPVT–R	3	0.43
Fischer et al., 1990	3	39	47	—	—	13.2	12.9	PPVT–R	3	0.88
	3	21	53	—	—	15.4	16.5	PPVT–R	3	0.79
Gansler et al., 1998	4 any	10	30	70	93	35.0	28.9	WAIS–R	2	0.24
Garcia-Sanchez et al., 1997	2 ADDH	35	16	80	69	14.9	14.7	WAIS	0	1.37
	2 ADDI	35	9	80	80	14.9	15.0	WAIS	0	1.83
Giedd et al., 1994	3	18	18	100	100	10.5	11.9	WISC–R	2	1.14
Gomez & Condon, 1999	4 C	15	15	80	73	9.9	9.3	WISC–III	1	0.39
Gorenstein et al., 1989	0	26	21	42	95	10.2	10.1	WISC–R	3	0.06
Grodzinsky & Diamond, 1992	3	30	34	100	100	7.5	7.6	WISC–R	1	0
	3	34	32	100	100	10.4	10.2	WISC–R	1	0.4
Halperin et al., 1992	3	18	31	9	77	9.2	9.6	PPVT–R	3	0.77
Herpertz et al., 2001	4 any	21	21	100	100	9.8	10.3	WISC–III	0	1.27
Holdnack et al., 1995	3	30	25	63	60	26.7	30.6	WAIS–R	1	1.18
Horn et al., 1989	3	19	37	100	100	8.2	8.1	PPVT–R	3	0.61
	3	12	17	0	0	7.9	8.2	PPVT–R	3	0.71
Houghton et al., 1999	4 C	28	62	54	53	10.2	10.4	WISC–III	2	0.4
	4 I	28	32	54	69	10.2	9.8	WISC–III	2	−0.1

(*table continues*)

Table 1 (*continued*)

Study	*DSM*[a]	Control *n*	ADHD *n*	Control: % male	ADHD: % male	Control age (years)	ADHD age (years)	IQ test	IQ estimated[b]	IQ *d*
Iaboni et al., 1995	3	17	19	100	100	10.5	10.8	WISC–R	1	0.32
Iaboni et al., 1997	3	18	18	100	100	11.0	10.6	WISC–R	1	0.5
Jacobsen et al., 1996	3	22	18	73	94	13.5	12.6	WISC–R	1	0.31
Janzen et al., 1995	3	8	8	100	100	11.1	10.5	WISC–R	0	0.6
Jennings et al., 1997	3	26	40	100	100	9.8	9.7	WISC–R	1	0.27
Johnson et al., 2001	4 any	38	56	63	71	40.8	33.3	Shipley	2	0.29
Jonkman et al., 1997	3	18	18	100	89	10.0	10.6	WISC–R	0	0.68
Jonkman et al., 1999	3	14	14	86	93	10.5	9.5	WISC–R	0	0.92
Jonkman et al., 2000	3	14	14	86	93	10.1	9.6	WISC–R	0	0.94
Kaplan et al., 1998	3	112	53	73	87	11.3	12.4	WISC–III	1	0.53
Kemner et al., 1995	2 ADDH	20	20	80	100	10.6	9.9	WISC–R	0	0.12
Kerns et al., 2001	4 C	21	21	76	76	9.3	9.4	K–BIT	1	0.54
Klorman et al., 1999	4 C, H	28	66	39	76	10.3	9.7	WISC–R	0	0.63
	4 I	28	51	39	67	10.3	9.6	WISC–R	0	1.05
Krane & Tannock, 2001	4 unsp	24	169	46	84	8.8	8.7	WISC–III	0	1.59
Kuperman et al., 1996	3	12	16	50	81	10.0	10.1	WISC–R	0	0.93
Lazzaro et al., 1999	4 C, I	54	54	100	100	13.4	13.7	K–BIT	1	0.66
Loge et al., 1990	3	20	20	85	85	9.5	9.6	WISC–R	0	0.9
Lorch et al., 2000	3	52	40	100	100	10.1	10.1	PPVT–R	3	0.25
	3	59	44	100	100	9.6	9.6	PPVT–R	3	0.41
Lufi, 2001	4 any	102	30	56	80	10.4	10.2	WISC–R	3	0.73
Mahone et al., 2001	4 any	28	21	43	67	10.8	11.7	WISC–III, WAIS–R	0	0.3
Mann et al., 1992	3	27	25	100	100	10.5	10.6	WISC–R	0	0.43
Mariani & Barkley, 1997	3	30	34	100	100	5.1	5.0	SB	1	0.57
Mataro et al., 1997	3	19	11	84	73	14.8	14.6	WAIS	0	0.99
Murphy et al., 2001	4 any	105	64	69	75	21.2	21.1	K–BIT	1	0.71
Newcorn et al., 1989	3	68	6	41	67	9.2	8.1	PPVT–R	3	−0.35
Nigg, 1999	4 C	25	25	100	68	10.1	9.6	WISC–III	1	0.34
Nigg et al., 1996	3	15	23	100	100	9.3	8.6	WISC–III	3	0.18
Nigg et al., 1997	3	17	27	100	100	9.3	9.9	WISC–R	3	0.83
Nigg et al., 1998	3	71	42	100	100	8.9	9.4	WISC–R, WISC–III	3	0.43
Nigg et al., 2002	4 C	21	22	75	83	21.6	23.1	WAIS–R	1	−0.26
Oie & Rund, 1999	3	30	20	53	100	14.1	15.7	WISC–R	1	1.15
Oosterlaan & Sergeant, 1996	0	17	15	41	87	8.7	9.3	WISC–R	1	0.58
Oosterlaan & Sergeant, 1998	0	21	14	71	86	10.1	10.1	WISC–R	1	1.21
Overtoom & Kenemans, 2002	3	16	16	88	100	10.4	10.3	WISC–R	0	0.7
Overtoom et al., 1998	4 C	16	16	88	100	10.3	10.4	WISC–R	0	0.7
Ozonoff & Jensen, 1999	4 unsp	29	24	—	—	12.1	11.1	WISC–III	0	0.06
Phelps, 1996	4 any	40	40	70	75	10.2	10.2	WISC–III	0	−0.26
Pisecco et al., 2001	2 unsp	281	20	100	100	11.0	11.0	WISC–R	0	0.47
Purvis & Tannock, 1997	3	14	14	100	100	9.2	8.7	WISC–R	1	0.37
Purvis & Tannock, 2000	3	17	17	65	94	9.5	9.1	WISC–III	0	−0.32
Ross et al., 1994	3	10	13	50	100	11.5	11.2	WISC–III	1	0.65
Rothenberger et al., 2000	3	11	11	100	100	11.5	11.1	—	—	0.45
Rovet & Hepworth, 2001	3	120	140	47	70	9.9	9.8	WISC–III	0	1.41
Rucklidge & Tannock, 2001	4 any	24	28	0	0	15.3	14.7	WISC–III	1	0.82
	4 any	20	35	100	100	14.8	14.8	WISC–III	1	0.77
Rund et al., 1998	3	30	20	53	100	15.7	14.1	WISC–R	2	0.72
Schachar & Logan, 1990	2 ADDH	10	13	—	—	10.0	9.3	WISC–R	1	1.06
Schachar & Tannock, 1995	3	16	22	100	100	9.0	9.2	WISC–R	1	0.13
Schachar et al., 1995	3	22	14	100	100	9.2	8.7	WISC–R	1	0.49
Schachar et al., 1988	2 ADDH	15	18	100	100	9.0	8.6	WISC–R	1	0.72
Schachar et al., 2000	4 any	33	72	61	80	9.3	9.0	—	—	1.08
Scheres et al., 2001	4 any	41	24	56	75	10.2	10.1	WISC–R	1	0.63
Schweitzer & Sulzer-Azaroff, 1995	3	8	10	100	100	6.3	6.1	WISC–R, SB–IV	1	0.27
Seidman et al., 1997	3	43	36	0	0	11.9	11.4	WISC–R	1	0.74
Seidman et al., 1997	3	99	118	100	100	15.3	14.5	WISC–R	1	0.5
Seidman et al., 1998	3	73	64	45	52	40.1	36.3	WAIS–R	1	0.12
Semrud-Clikeman et al., 1996	4 C	10	10	80	80	11.8	10.0	WISC–R	0	1.1

Table 1 (*continued*)

Study	*DSM*[a]	Control *n*	ADHD *n*	Control: % male	ADHD: % male	Control age (years)	ADHD age (years)	IQ test	IQ estimated[b]	IQ *d*
Shapiro et al., 1993	3	38	67	74	84	9.0	8.9	K–BIT	1	0.04
Silberstein et al., 1998	3	17	17	100	100	11.0	10.8	WISC–III	1	0.87
Smith et al., 2002	4 C, H	22	22	91	86	11.2	11.3	WISC–III	1	0.64
Steger et al., 2000	3	16	15	100	100	10.8	10.8	WISC–R	0	0.74
Steger et al., 2001	3	20	22	85	86	10.6	10.9	WISC	0	0.46
Stevens et al., 2002	0	76	76	76	68	10.0	10.0	K–BIT	1	0.88
Stewart et al., 2001	4 C	9	9	100	100	11.7	10.9	WISC–III	0	0.09
	4 I	9	9	100	100	11.7	11.6	WISC–III	0	0.51
Swaab-Barneveld et al., 2000	3	55	52	100	100	9.8	9.4	WISC–R	1	0.6
Tannock et al., 2000	4 unsp	27	67	70	87	9.1	8.9	WISC–III	2	0.16
Tarnowski et al., 1986	2 ADDH	13	14	100	100	8.6	8.4	WISC–R	0	0.73
Tripp & Alsop, 2001	4 C	36	36	86	92	8.3	8.7	WISC–III, WPPSI–R	0	1.03
Walker et al., 2000	4 unsp	30	30	67	83	25.8	25.8	WAIS–R	2	1.03
Weiler et al., 2000	4 I	24	16	64	88	9.5	9.5	WISC–III	0	0.37
West et al., 2000	4 C	44	30	100	100	10.1	9.9	WISC–III	2	−0.16
	4 I	44	14	100	100	10.1	10.7	WISC–III	2	−0.25
Weyandt et al., 2002	5	62	17	34	59	—	—	WAIS–R	0	−0.56
Weyandt & Willis, 1994	3	45	36	53	100	9.0	9.2	PPVT–R	3	−0.22
Wiers et al., 1998	4 C	34	28	100	100	9.1	9.0	WISC–R	1	0.77
Willcutt et al., 2001	2 unsp	121	52	—	—	10.7	10.8	WISC–R	0	1.21
Wu et al., 2002	4	29	58	—	—	10.6	10.5	WISC–III	1	0.81
Zametkin et al., 1990	2 ADDH	50	25	56	72	36.3	37.4	WAIS–R	1	0.55
Zametkin et al., 1993	3	10	10	70	70	14.3	14.5	WISC–R	1	0.41

Note. Full citations for references used in the meta-analysis are available on the Web at http://dx.doi.org/10.1037/0894-4105.18.3.543.supp. *DSM* = *Diagnostic and Statistical Manual of Mental Disorders;* FSIQ = Full Scale IQ; WISC = Wechsler Adult Intelligence Scale; WISC–R = Wechsler Intelligence Scale for Children—Revised; WISC–III, Wechsler Intelligence Scale for Children—Third Edition; WAIS = Wechsler Adult Intelligence Scale; WAIS–R = Wechsler Adult Intelligence Scale—Revised; Shipley = Shipley Institute of Living Scale, total scale; SB = Stanford–Binet Intelligence Scale, unspecified edition; SB–IV = Stanford–Binet Intelligence Scale—Fourth Edition; K–BIT = Kaufmann Brief Intelligence Test; PPVT–R = Peabody Picture Vocabulary Test—Revised; WASI = Wechsler Abbreviated Scale of Intelligence; WPPSI–R = Wechsler Primary and Preschool Scale of Intelligence—Revised; IQ *d* = FSIQ effect size. Dash indicates data were not reported.
[a]0 = no criteria reported; 2 = *DSM–III* criteria; 3 = *DSM–III–R* criteria; 4 = *DSM–IV* criteria; C = combined type; H = hyperactive–impulsive type; I = inattentive type; any = any subtype; unsp = subtype unspecified; ADDH = *DSM–III* criteria for attention-deficit disorder, hyperactive type; ADDI = *DSM–III* criteria for attention-deficit disorder, inattentive type. [b]0 = not estimated; 1 = estimated from two or more subtests; 2 = estimated by averaging the effect sizes derived from two or more subtests or scales; 3 = estimated from single subtest.

typically best analyzed by hierarchical linear modeling, the small number of effect sizes contributed by these studies ($n = 27$) relative to the total number of effect sizes ($n = 137$) suggested that little additional information would be gained from these analyses (Bryk & Raudenbush, 1992). Therefore, effect sizes were considered independent in the following analyses, and subsequent *n*s indicate the number of comparisons, not the number of studies. Comparisons in which FSIQ was estimated by the authors or could be estimated from the data presented ($n = 86$) were included, along with comparisons in which FSIQ was estimated from the complete test ($n = 47$) and comparisons in which the estimation procedure could not be determined ($n = 4$). For comparisons in which FSIQ was estimated from incomplete tests, 54 used a standard short form such as 2 or more subtest short forms of the Wechsler scales; 21 used data from a single Wechsler scale subtest, usually Vocabulary, Block Design, or Similarities; and 11 were estimated by combining the effect sizes derived from two or more subtests. Of the 137 comparisons for FSIQ, 86 also had data from at least one other neuropsychological test. Data from neuropsychological tests with fewer than six comparisons were excluded from the present review.

Only studies including data from a healthy comparison group without significant pathology were included in the present meta-analysis. Studies in which the comparison group contained some participants without ADHD but with other conditions (e.g., oppositional defiant disorder, conduct disorder, learning disability) were excluded, as were studies in which conditions other than ADHD were deemed the primary problem of the clinical group.

Dependent Variables

The following is a list of measures examined in the present review, with measures separated into intellectual composites, achievement measures, nonexecutive functioning measures, and executive functioning measures. The latter distinctions were made to clarify which measures were expected to be more and less sensitive to ADHD. Although all cognitive measures are likely affected by executive dysfunction to some extent, measures classified as executive functioning indicators were expected to be significantly more sensitive than were nonexecutive measures.

Intellectual composites. In the articles we reviewed, FSIQ was estimated primarily from the Wechsler scales (Wechsler, 1974, 1981, 1989, 1991, 1999). However, some studies estimated FSIQ using the Kaufman Brief Intelligence Test (Kaufman & Kaufman, 1990), the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1981); the Stanford–Binet Intelligence Scale (Thorndike, Hagen, & Sattler, 1986), or the Shipley Institute of Living Scale (Shipley, 1940). Verbal IQ and Performance IQ are both derived from the Wechsler scales. Verbal IQ estimates verbal reasoning and language abilities, and Performance IQ estimates visual perceptual and visual reasoning abilities. Freedom from Distractibility is an aggregate measure that combines Arithmetic and Digit Span. This index was obtained from studies using the *Wechsler Intelligence Scale for Children—Revised* (1974), the *Wechsler Intelligence Scale for Children—Third Edition* (WISC–III; Wechsler, 1991), and the *Wechsler Adult Intelligence Scale—Revised* (1981). Processing speed is also an aggregate measure that combines Digit Symbol–Coding and Symbol Search from the WISC–III.

Achievement. The Wide Range Achievement Test—Revised and *the Wide Range Achievement Test—Third Edition* (WRAT; Jastak & Wilkinson, 1984; Wilkinson, 1993) consist of three subtests measuring reading, spelling, and math skills. These subtests are used to identify individuals with learning disabilities or to identify problems underlying some academic deficits, such as phoneme to grapheme and grapheme to phoneme conversion.

Nonexecutive measures. The Vocabulary subtest from the Wechsler scales measures the ability to give word definitions. Block Design measures visual constructional abilities, and the Similarities subtest assesses the ability to find the abstract quality shared by two words. The PPVT measures receptive language ability and has shown high correlations with measures of FSIQ (Dunn & Dunn, 1981). The Rey Complex Figure (Rey & Osterrieth, 1993) is an abstract visual design that is difficult to encode verbally. The test consists of having to copy the design as accurately as possible and then without warning having to copy the design from memory after a 20- to 45-min delay. The measures of interest in the present review were accuracy scores for the copy trial (Rey Copy) and recall trial (Rey Recall). Rey Copy largely reflects visual constructional abilities, and Rey Recall measures long-term visual memory.

Executive functioning measures. The Digit Span subtest from the Wechsler scales assesses the ability to hold a series of digits in working memory and loads heavily on focused attention and working memory. Digit Symbol–Coding measures the ability to copy symbols quickly and accurately from a key and loads heavily on focused attention and processing speed. Above-average performance on this subtest also requires working memory ability (Kaufman, 1990). Arithmetic measures the ability to do relatively simple math problems in memory and loads heavily on working memory.

Continuous performance tests (CPTs) measure the ability to maintain focused attention over longer time periods (usually 5–20 min) while responding to target stimuli and inhibiting responses to nontarget stimuli. Several different versions of CPTs were used in the studies reviewed, including the CPT from the Gordon Diagnostic System (Gordon, 1983), the Conners CPT (Conners, 1994), and several unspecified research versions. The dependent measures of interest for the present review were the number of hits recorded (CPT–Hits), the mean reaction time (RT) to targets (CPT–MRT), the number of omission errors (CPT–Omission), and the number of incorrect responses to nontargets (CPT–Commission). CPT–Hits measures overall target and nontarget discrimination, CPT–MRT reflects visual processing speed, CPT–Omission reflects sustained attention or vigilance to the task, and CPT–Commission reflects behavioral inhibition.

The Stop Signal Task (SST) was developed to measure behavioral inhibition. Several paradigms have been reported in the literature, but each reflects similar task characteristics in which an individual is asked to respond quickly to a RT task but to inhibit that response occasionally and sporadically when a stop signal stimulus (visual or auditory) precedes the target. Three measures of interest were examined in the present review—RT to go-target stimuli (SST–Go), RT to stop signals (SS–Stop), and the probability of inhibition (SST–Probability of Inhibition; for a complete description, see Oosterlaan, Logan, & Sergeant, 1998).

Trail Making Test—A and Trail Making Test—B (Lezak, 1995; Reitan, 1958) measure the ability to draw lines between numbers in numerical order (Trail–A) and alternate between numbers and letters in numerical and alphabetical order (Trail–B). Trail–A is a measure of focused attention, and Trail–B loads heavily on focused attention and working memory.

The Wisconsin Card Sorting Task (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993) is a commonly used measure of executive functioning and requires an individual to match a series of cards on one of three stimulus characteristics without being instructed regarding how the cards should be matched. The only feedback given is whether a match is correct or incorrect. The measures of interest were the number of perseverative responses or errors (WCST–Perseveration), categories completed (WCST–Categories), and the number of times that an individual lost set (WCST–Set Failures). WCST variables reflect a number of executive functions, including working memory, behavioral inhibition, and set shifting.

The Stroop Interference Task (Stroop, 1935) consists of naming the color of ink in which a word is printed instead of reading the color word and reflects behavioral inhibition. The dependent measure was calculated slightly differently across studies but was generally derived from the number of trials completed or the time required to complete the task.

The Matching Familiar Figures Test (MFFT; Kagan, 1966) is a task in which children are shown a page with a picture at the top and six similar pictures below. Children are required to pick out the picture from the six choices that is identical to the sample picture. The measures of interest for the present review were the latency to first response (MFFT–Time) and the number of response errors (MFFT–Errors). Both measures are thought to reflect behavioral inhibition.

The Word Fluency Test (Lezak, 1995) consists of having subjects produce as many words as possible within 1 min that start with a letter of the alphabet (Letter Fluency) or that comprise a particular semantic category (Category Fluency). Both of these measures are thought to be sensitive to deficits in expressive-language abilities (Benton & Hamsher, 1976) and are often considered to be sensitive to frontal-lobe deficits (for a discussion, see Lezak, 1995). However, because these measures are sometimes considered to be language and not executive function measures (Mitrushina et al., 1999; Spreen & Strauss, 1998), analyses were performed with and without these measures.

Evaluation of methodological factors. To determine the influence of methodological factors on the magnitude of ADHD and control participant differences on intellectual and neuropsychological tests, we coded each comparison for several variables, including medication status (meds; yes = not on meds, no = on meds, or medication status not reported), IQ test reported (IQ; yes = 1, no = 0), age of groups reported (age), impairment in at least two settings reported as a criterion in making diagnosis (impairment), structured or semistructured diagnostic interview versus no interview used in making the diagnosis (interview), standardized rating scale used in making diagnosis (rating), empirically derived standardized rating scale cutoffs used in making diagnosis (cutoff), multiple sources of information required for diagnosis (information), oppositional defiant disorder–conduct disorder comorbidity excluded, learning disability comorbidity excluded, neurological and physical or sensorimotor problems ruled out (physical), referral sources for control and ADHD the same or different (referral; 1 = same, 0 = different), control participants screened for psychiatric diagnoses by using a structured or semistructured interview or comprehensive rating scales (screen), and gender composition of sample reported (gender). Coding was performed by Thomas W. Frazier for all studies. This procedure was used for two reasons. First, although the coder was not blind to study hypotheses, each rating consisted of simple presence and absence judgments. Second, the author had no prior contact with any of the authors of studies included in the present meta-analysis, an important source of bias (Rosenthal, 1991). To examine any problems with rater reliability, another rater randomly coded 40 comparisons on all methodological characteristics. For these studies, agreement was perfect for all variables, except for two ratings for the impairment criterion and one for the interview criterion, which were easily resolved.

Five additional variables were coded for each comparison, and these variables were used in separate analyses to determine their influence on ADHD and control participant differences in cognitive abilities. The diagnostic scheme used to classify ADHD participants was coded into four categories: *DSM–IV* (American Psychiatric Association, 1994; $n = 53$), *DSM–III–R* (American Psychiatric Association, 1987; $n = 67$), *DSM–III* (American Psychiatric Association, 1980; $n = 13$), and no diagnostic scheme ($n = 4$). The type of test used to derive FSIQ was coded into three categories: Wechsler scales ($n = 108$), PPVT IQ estimates ($n = 15$), and other intellectual assessment batteries ($n = 14$). The procedure used to estimate FSIQ was coded into four categories on the basis of the previously described methodology. Average age was determined by averaging the ages of control and ADHD participants, weighted by sample sizes. Differences in the gender composition of ADHD and control samples were computed by subtracting the percentage of male participants in the control group from the percentage in the ADHD group.

Calculation of effect sizes. For each measure, we computed the standardized mean difference in effect size using pooled standard deviations, Cohen's d, with formulas provided in Lipsey and Wilson (2001). For each measure, the unweighted mean, median, and weighted mean effect sizes were also calculated along with the 95% confidence interval for the weighted mean effect size. Homogeneity analyses were also computed for each measure. The latter test evaluates whether the observed effect sizes likely result from sampling one population of effect sizes. For measures with heterogeneous distributions of effect sizes, correlations between methodological factors and effect sizes were performed to determine whether methodological characteristics were related to the magnitude of differences between ADHD and control groups.

Results

ADHD Versus Control Participants: Overall Cognitive Ability

Table 2 presents the unweighted and weighted mean effect sizes, median effect sizes, the confidence interval for the weighted mean effect size, and the Q statistic for the homogeneity of effect sizes for each measure. For FSIQ, less than half of the comparisons (63 of 137) represented statistically significant findings at the .05 level of significance, and one of the significant findings was opposite of the predicted direction, with ADHD participants showing higher FSIQ than control participants. To determine whether ADHD participants generally score lower on measures of overall cognitive ability than do healthy control participants (Hypothesis 1), we computed the weighted mean effects size d and the z test for the

Table 2
Mean, Median, and Weighted Mean Effect Sizes and the 95% Confidence Intervals (CIs) and Q Statistics

Measure	n	M	Mdn	Weighted M	95% CI	Q
FSIQ	137	.60	.60	.61	.57–.65	353.49**
Verbal IQ	29	.59	.53	.67	.58–.76	136.56**
Performance IQ	26	.55	.36	.58	.48–.68	116.94**
PPVT	15	.41	.41	.41	.28–.54	27.29*
Vocabulary	22	.50	.48	.48	.39–.57	61.01**
Block Design	25	.41	.46	.44	.35–.53	58.08**
Digit Span	12	.66	.70	.64	.52–.76	25.30**
Digit Symbol–Coding	15	.85	.95	.82	.72–.92	62.44**
Similarities	6	.62	.55	.64	.40–.88	5.77
Arithmetic	9	.72	.72	.70	.57–.83	15.60*
Freedom from Distractibility	12	.67	.77	.75	.62–.88	40.20**
Processing Speed	6	.80	.85	.65	.42–.88	14.92**
CPT–MRT	17	.48	.43	.39	.26–.52	38.63**
CPT–Hits	19	.91	.75	1.00	.87–1.13	97.16**
CPT–Omission	33	.72	.72	.66	.58–.74	52.75**
CPT–Commission	40	.59	.59	.55	.47–.63	61.93**
SST–Go RT	10	.71	.69	.66	.46–.86	14.66
SST–Stop RT	13	.62	.57	.54	.40–.68	6.41
SST–Probability of Inhibition	9	.40	.54	.34	.18–.50	16.88*
Trail–A	13	.45	.47	.40	.26–.54	18.33
Trail–B	14	.64	.59	.59	.46–.72	17.42
WCST–Perseveration	25	.36	.39	.35	.26–.44	23.13
WCST–Categories	22	.32	.28	.29	.19–.39	23.78
WCST–Set Failure	14	.18	.17	.15	.03–.27	14.24
Stroop Interference	20	.63	.54	.56	.46–.66	41.81**
MFFT–Time	11	.28	.30	.27	.12–.42	10.18
MFFT–Errors	10	.64	.59	.60	.44–.76	7.61
Rey Copy	6	.29	.28	.24	.07–.41	3.69
Rey Recall	6	.44	.38	.26	.09–.43	10.24
Letter Fluency	13	.54	.36	.46	.33–.59	40.45**
Category Fluency	9	.46	.43	.41	.24–.58	9.94
WRAT–Reading	24	.70	.93	.64	.53–.75	109.70**
WRAT–Spelling	15	.88	1.08	.87	.72–1.02	27.04**
WRAT–Arithmetic	21	1.01	.89	.89	.78–1.00	74.82**

Note. FSIQ = Full Scale IQ; PPVT = Peabody Picture Vocabulary Test; CPT = continuous performance test; MRT = mean reaction time; SST = Stop Signal Task; RT = reaction time; WCST = Wisconsin Card Sorting Test; WRAT = Wide Range Achievement Test; MFFT = Matching Familiar Figures Test. n = number of effect sizes; Mdn = Median; Weighted M = weighted mean effect size; Q = Q statistic distributed as χ^2 with $n - 1$ degrees of freedom.
$^*p < .05$. $^{**}p < .01$.

significance of this effect size (Lipsey & Wilson, 2001). ADHD groups displayed significantly lower FSIQ scores relative to control groups (weighted mean $d = .61$, $z = 27.72$). The significance of the weighted mean effect size is unlikely to be a result of publication bias because 140 comparisons with null results would be needed to bring the mean effect size down to $d = .30$, a small effect, and more than 1,500 studies with null results would be needed to reduce the mean effect size to $d = .05$, a negligible effect (Orwin, 1983). Studies including multiple effect sizes did not differ on FSIQ from studies including only one comparison, $t(135) = 0.57$, $p = .57$, further suggesting that treating multiple effect sizes from a single study as independent did not substantially influence findings for FSIQ. Verbal and Performance IQs were also significantly sensitive to ADHD (Verbal IQ, $d = .67$, $z = 14.13$; Performance IQ, $d = .58$, $z = 11.71$).

To determine whether ADHD subtypes differed on FSIQ, within-study effect sizes were computed for hyperactive–impulsive and combined types and inattentive-type groups. A paired-samples *t* test was used to examine subtype differences. ADHD subtypes did not significantly differ on FSIQ (inattentive subtype, $d = .78$; hyperactive–impulsive and combined subtypes, $d = .58$); $t(7) = 1.44$, $p = .19$.[1] However, because of the small number of comparisons, these results may not reflect true equivalence. Therefore, all analyses and effect sizes were recomputed without ADHD–inattentive comparisons. As expected given recent findings regarding subgroup differences, analyses and effect sizes did not change substantially by excluding comparisons with inattentive subjects (largest change in effect size, $\Delta d = .08$). All comparisons were included in subsequent analyses.

Neuropsychological Measures

To examine the hypothesis that all neuropsychological measures would be sensitive to ADHD (Hypothesis 2), we determined the weighted mean effect sizes and *z* tests for the significance of the weighted mean effect sizes in the same fashion as described above for FSIQ. All measures were significantly sensitive to ADHD (smallest $z = 2.5$, $p < .05$). However, there was wide variability in the effect sizes observed, ranging from a very small effect size for WCST–Set Failure (.15) to a very large effect size for CPT–Hits (1.0). Only 28 comparisons finding zero difference between ADHD and control participants would be needed to reduce the mean effect size for WCST–Set Failure to $d = .05$, a negligible effect, whereas 361 comparisons with zero difference would be needed to reduce the mean effect size for CPT–Hits to $d = .05$.

FSIQ Versus Neuropsychological Measures

To examine differences between effect sizes for FSIQ and neuropsychological measures (Hypothesis 3), we computed paired-samples *t* tests between pairs of weighted effect sizes for FSIQ and the weighted effect sizes for other cognitive measures. Table 3 presents the number of comparisons included in these tests and the test statistic. Effects sizes for FSIQ were significantly larger than those for Block Design, $t(24) = 3.90$, $p < .001$; SST–Probability of Inhibition, $t(8) = 2.75$, $p = .025$; WCST–Categories, $t(21) = 2.71$, $p = .01$; WCST–Set Failures, $t(13) = 2.18$, $p = .048$; MFFT–Time, $t(10) = 2.25$, $p = .048$; Rey Copy, $t(5) = 2.57$, $p = .05$; and Rey Recall, $t(5) = 4.412$, $p < .01$, and were marginally larger than those for WCST–Perseveration, $t(24) = 1.86$, $p = .08$. Effect sizes for FSIQ were also significantly smaller than those for WRAT–Spelling, $t(13) = -3.41$, $p < .01$, and WRAT–Arithmetic, $t(20) = -3.23$, $p < .01$, and marginally significantly smaller than those for CPT–Hits, $t(18) = -1.93$, $p = .07$, and CPT–Omission, $t(32) = -1.96$, $p = .06$. No other significant or marginally significant differences were present between FSIQ and other cognitive measures.

Table 3
Number of Paired Effect Sizes (ESs) and Test Statistics for Each FSIQ Versus Neuropsychological Test Comparison

Measure vs. FSIQ	ES	*t*	*p*
Verbal IQ	29	1.22	.23
Performance IQ	26	1.71	.10
Vocabulary	22	0.98	.34
Block Design	25	3.90	.00
Digit Span	12	0.16	.87
Digit Symbol–Coding	15	−0.90	.38
Similarities	6	1.37	.23
Arithmetic	9	−0.05	.96
Freedom from Distractibility	12	−1.59	.14
Processing Speed	6	−0.92	.40
CPT–MRT	17	0.54	.60
CPT–Hits	19	−1.93	.07
CPT–Omission	33	−1.96	.06
CPT–Commission	40	−0.27	.79
SST–Go RT	10	−0.45	.66
SST–Stop RT	13	0.21	.84
SST–Probability of Inhibition	9	2.75	.03
Trail–A	13	0.57	.58
Trail–B	14	−0.97	.35
WCST–Perseveration	25	1.86	.08
WCST–Categories	22	2.71	.01
WCST–Set Failure	14	2.18	.05
Stroop Interference	19	−1.18	.25
MFFT–Time	11	2.25	.05
MFFT–Errors	10	−1.30	.22
Rey Copy	6	2.57	.05
Rey Recall	6	4.41	.01
Letter Fluency	13	−0.03	.98
Category Fluency	9	−0.30	.77
WRAT–Reading	23	−0.98	.34
WRAT–Spelling	14	−3.41	<.00
WRAT–Arithmetic	21	−3.23	<.00

Note. FSIQ = Full Scale IQ; CPT = continuous performance test; MRT = mean reaction time; SST = Stop Signal Task; RT = reaction time; WCST = Wisconsin Card Sorting Test; MFFT = Matching Familiar Figures Test; WRAT = Wide Range Achievement Test.

Spared Versus Affected Neuropsychological Functions

To determine whether executive functioning measures would generally display larger effect sizes than measures of nonexecutive functions (Hypothesis 4), we averaged effect sizes for measures of visual constructional (Block Design and Rey Copy), visual memory (Rey Recall), and receptive and expressive language (PPVT Vocabulary, Similarities) within each study. This was done to

[1] For this analysis, data from the Chhabildas et al. (2001) study for hyperactive–impulsive and combined subtype groups were weighted by sample size and then averaged.

create one aggregate variable reflecting cognitive functions that were thought to be relatively spared in ADHD. A separate aggregate variable was also computed for each study by averaging the effects sizes for measures thought to be most affected in ADHD, including measures of behavioral inhibition (WCST–Perseveration, WCST–Set Failure, WCST–Categories, SST–Stop, SST–Probability of Inhibition, Stroop Interference, MFFT–Time, MFFT–Errors, and CPT–Commission), focused and sustained attention (Digit Symbol–Coding, Symbol Search, CPT–MRT, CPT–Hits, CPT–Omission, SST–Go, Trail–A), working memory (Digit Span, Arithmetic, Trail–B), and verbal fluency (Letter Fluency and Category Fluency). If only a single measure was present in a given study for estimating spared or affected cognitive functions, the effect size for that measure was used as the estimate (Rosenthal & Rubin, 1986). A paired-samples *t* test was used to compare estimates of spared and affected functions in ADHD. Twenty-nine studies included effect sizes for each variable. As hypothesized, measures thought to assess nonexecutive cognitive abilities in ADHD produced a significantly smaller weighted mean effect size ($d = .39$, $SE = .06$) compared with measures of cognitive functions that were impaired in ADHD ($d = .58$, $SE = .05$), $t(28) = 2.92$, $p = .007$. Excluding fluency measures did not substantially change these findings, $t(27) = 2.74$, $p = .01$, nor did including fluency measures as nonexecutive measures, $t(31) = 3.45$, $p = .002$.

Methodological Factors

Two approaches were used to determine the influence of methodological factors on the magnitude of effect sizes of cognitive measures (methodological characteristics of each study can be found in the table on the Web at http://dx.doi.org/10.1037/0894-4105.18.3.543.supp). The first approach involved examining the relationship between empirically derived methodological factors and effect sizes comparing ADHD and control groups. In this approach, the 14 methodological factors were subjected to principal-components analysis. Although dichotomous variables are not optimal input for principal-components analysis, the present application was used for the heuristic purpose of understanding the methodological factors responsible for within-measure variability in effect sizes (Kim & Meuller, 1978). The present analytic strategy is likely superior to computing correlations between individual variables because the latter approach would significantly inflate Type 1 error rates. Component-based scores will also be more reliable than their constituent item scores, improving statistical power. Horn's parallel analysis and minimum average partial analysis were used as the criterion for determining the number of components to retain because these criteria have been shown to yield the most accurate number of components in Monte Carlo studies (Zwick & Velicer, 1986). Both criteria indicated a five-component solution, with the five factors accounting for 57% of variance in the original scores. An oblique rotation, oblimin, was used to determine which variables loaded above .50 on the components.

The first factor, labeled *Multiple Sources,* consisted of the highest positive loadings from information (.82) and impairment (.80). This factor represents whether additional sources of information, usually obtained from multiple informants, were used in making the ADHD diagnosis. The second factor, labeled *Clinician Diagnosis,* consisted of a high positive loading from interview (.61) and high negative loadings from rating scale (−.64) and cutoff (−.62). Studies scoring high on this factor relied primarily on information generated through clinical interview, whereas those scoring low used information from rating scales for making clinical diagnoses. The third factor, labeled *Reporting,* had high loadings from IQ (.83) and controls screened (.77). This factor likely represented thoroughness in reporting procedures because only 7% of studies did not report screening control participants for potential psychiatric illnesses, and 4% did not report the IQ test given. The fourth factor, labeled *Screening Comorbidity,* consisted of high loadings from LD (.62) and ODD/CD (.77). Studies high in this factor attempted to screen for comorbid psychiatric conditions. The fifth factor, labeled *Clinical/Community Sample,* consisted of a high positive loading from referral (.70) and high negative loadings from physical problems (−.58) and age (−.49). High scores on this factor indicated use of community versus clinical samples, because studies using clinical samples frequently used different referral sources to obtain control participants and patients. Studies using community samples typically did not report the ages of participants and did not screen for physical or neurologic problems because participants were frequently from the same grades and were assumed to be in good health.

FSIQ effect sizes showed significant correlations with multiple sources ($r = .18$, $p = .04$) and clinician diagnosis ($r = .17$, $p = .05$) and were significantly inversely related to clinical/community sample ($r = -.20$, $p = .02$). No significant relationships were found for Reporting and Screening Comorbidity. Only 6 of 165 correlations were significant between methodological factors and effect sizes for other cognitive and neuropsychological measures. These findings were likely due to the number of statistical tests; therefore, specific relationships are not reported.

The second approach involved examining the effect of the remaining methodological variables (type of intelligence measure, method for estimating overall ability, diagnostic scheme, publication date, average age of participants, and gender differences in the composition of the control and ADHD samples) on the effect sizes for FSIQ. Effect sizes for FSIQ did not differ as a function of the type of intelligence test used, $F(2, 133) = 1.71$, $p = .19$, diagnostic scheme used to classify participants, $F(3, 132) = 0.58$, $p = .63$, or publication date ($r = .11$, $p = .19$). However, effect sizes for FSIQ did differ as a function of the way in which FSIQ was estimated. Comparisons in which FSIQ was estimated using complete tests showed the largest unweighted mean effect sizes ($d = .73$, $SE = .07$) and were significantly different from comparisons in which FSIQ was estimated by using effect sizes from multiple subtests ($d = .38$, $SE = .13$; contrast $p = .01$) and comparisons in which FSIQ was estimated using a single subtest ($d = .45$, $SE = .09$; contrast $p = .01$). The latter two estimation procedures did not significantly differ from each other (contrast $p = .97$), and estimation using short forms did not significantly differ from any of the other estimation techniques ($d = .56$, $SE = .05$; smallest $p = .22$). Effect sizes for FSIQ were unrelated to the average age of participants ($r = -.01$, $p > .90$) or to gender differences in the composition of the sample ($r = .01$, $p > .90$).

Discussion

The hypothesis that ADHD and control participants would significantly differ in overall cognitive ability was supported. The weighted mean effect size (d = .61) is roughly equivalent to a 9-point difference in FSIQ for most commercial IQ tests. This finding is consistent with research positing medium to large differences in overall ability (Barkley, DuPaul, & McMurray, 1990) and may indicate that the disorder is characterized by mild global cognitive inefficiencies or by multiple specific deficits affecting several cognitive abilities. It is also possible that a decrement of this magnitude on FSIQ is accounted for by test-taking differences between groups (Glutting, Youngstrom, Oakland, & Watkins, 1996). FSIQ was not significantly different between ADHD subtypes, suggesting that measures of overall ability are similarly decreased in the two major ADHD subgroups. However, the small number of studies including inattentive groups significantly reduced the power of these comparisons. Future studies are needed to further specify subtype differences in general ability and specific neuropsychological processes.

As expected, effect sizes for all neuropsychological measures were significantly larger than zero. However, only a few of the effect sizes for executive functioning measures emerged as significantly larger than effect sizes for FSIQ. In particular, visual constructional and visual memory measures were smaller, as predicted, but verbal fluency and language measures did not significantly differ from FSIQ. Unexpectedly, several executive functioning measures were also found to have significantly smaller effect sizes than FSIQ, including WCST–Variables, SST–Probability of Inhibition, and MFFT–Time. Only academic achievement and CPT measures displayed substantially larger effects than FSIQ, with effect sizes for several measures of attention, working memory, and behavioral inhibition being either equivalent or significantly smaller than FSIQ. Overall, this pattern of findings may suggest that not all aspects of executive functioning are equally impaired in ADHD and that other nonexecutive deficits may be present.

Despite the lack of consistent executive deficits when comparing individual measures with FSIQ, the analysis directly comparing the mean effect sizes of affected versus spared neuropsychological measures provided support for the notion that measures of executive functions are generally more impaired than are measures of primarily nonexecutive functions. Specifically, this finding provided limited empirical support for the behavioral inhibition and attention and working memory components of Barkley's (1997a) unified model of ADHD, because the measures included in the *affected* category load heavily on these dimensions. Additional studies are needed to test the self-regulation of affect, internalization of speech, reconstitution, and motor control components of this model.

Although the present findings do not support the notion of a generalized executive deficit, the possibility that a more specific impairment of executive functions accounts for differences in overall ability should be acknowledged. This possibility is highlighted by the fact that measures of overall ability are heavily influenced by executive functions (Barkley, 1997a; Mariani & Barkley, 1997). For example, several measures that are commonly included in estimates of overall ability assess working memory and processing speed. Executive functions may also influence estimates of overall ability through alterations in response style, approach to cognitive testing, affect regulation during the evaluation, and problems with response control required by most measures. Future studies are needed to sort out the contributions of these factors and executive functions to omnibus ability estimates as well as the influence of executive deficits on specific ability measures. As a result of the observed differences in overall ability between ADHD and control participants, several studies have attempted to control for existing differences in FSIQ when examining specific neuropsychological deficits (Barkley, Murphy, & Bush, 2001; Murphy, Barkley, & Bush, 2001). The present findings indicate that this approach is methodologically tenuous, because decrements in overall ability are a feature of the disorder, making statistical "control" impossible (Campbell & Kenny, 1999). Miller and Chapman (2001) recommended an alternative approach that involves determining the pattern of impaired cognitive functions in ADHD by using comparison samples with similar decrements in overall ability, such as individuals with seizure disorders, mild head trauma, or psychiatric difficulties.

The present review also revealed significant within-measure heterogeneity of effect sizes for most measures. However, the relationships between methodological factors and effect sizes were generally modest. For FSIQ, effect sizes were largest in situations in which ability was estimated from complete tests. This may be due to the heavy use of less sensitive, expressive, and receptive language measures (e.g., Vocabulary subtest and PPVT) in short form or single-test estimation procedures. Additionally, subtest short forms tend to have lower reliability than do FSIQ estimates that are based upon an entire battery, attenuating effect sizes for these comparisons (Spearman-Brown prophecy; see Glutting, Watkins, & Youngstrom, 2003). The smaller effect sizes for FSIQ when estimated from single subtests may also be due to longer testing times in studies using a full intellectual assessment battery. In the latter case, decreasing performance over time, caused by deficient sustained attention, would be expected in ADHD participants regardless of the specific functions assessed.

Effect sizes for FSIQ were also larger in studies using clinical samples and interview-based diagnoses and in which multiple sources of information were included in diagnostic decisions. This is likely caused by the identification of more severely disabled ADHD participants through the application of stricter diagnostic criteria (i.e., conjunctive combination of information; Youngstrom, Findling, & Calabrese, 2003). It is interesting that the diagnostic scheme used did not influence the magnitude of the effect, suggesting that the definitional changes across *DSM–III*, *DSM–III–R*, and *DSM–IV* have not had a substantial influence on the validity coefficients for FSIQ. Similarly, the age of participants did not influence effects for FSIQ. Although specific hypotheses were not made regarding age, this finding runs counter to behavioral descriptions of adult ADHD that imply improvements in hyperactivity and disinhibition. However, this may be due to the selection of adults who are still experiencing significant symptoms (Campbell & Kenny, 1999). The present data suggest that, at least in individuals with enduring behavioral symptoms, cognitive deficits persist into adulthood.

Although several methodological factors influenced the effect sizes for FSIQ, few significant relationships were observed for specific intellectual and neuropsychological measures. This is likely due to the small number of comparisons available, limiting the power of these tests. However, the lack of significant relations may suggest that additional unanalyzed factors contributed to the observed heterogeneity. Specifically, other unidentified sample characteristics, idiosyncratic test versions, and differences in test administration may all have contributed to the wide range of within-measure effect sizes. Differences in the reliability of neuropsychological tests may also contribute to the unclear pattern of observed effect sizes, further blurring the nature of cognitive deficits in ADHD. Previous meta-analyses have determined the factors responsible for the observed heterogeneity of effect sizes in ADHD on the CPT and the SST (Losier, McGrath, & Klein, 1996; Oosterlaan, Logan, & Sergeant, 1998). However, future meta-analytic reviews are needed delineating the factors responsible for heterogeneity across other measures, such as phonemic fluency measures, the Stroop Interference Task, and the WRAT subtests.

Another important finding of the present review was that academic measures of spelling and arithmetic skills were significantly more sensitive to ADHD than overall cognitive ability. One explanation for this observation is that few of the studies reviewed attempted to screen for learning disabilities in math or written expression. High comorbidity rates have been found between ADHD and learning disabilities (Barkley, 1998), and comorbid individuals show the largest deficits in executive functioning (Seidman, Biederman, Monuteaux, Doyle, & Faraone, 2001). The latter finding suggests a link between executive functioning and performance on achievement measures. Consistent with this interpretation, measures of academic achievement may be sensitive to subtle deficits in neurocognitive abilities as well as the behavior problems observed in ADHD that inhibit learning (Slomka, 1998). In this scenario, achievement measures may be useful not only for screening comorbid learning disabilities but also for characterizing behavioral and motivational deficits resulting from executive dysfunction. Future work should examine the structural relationships between particular types of executive dysfunction, behavioral symptoms of ADHD, and achievement. This work will determine the contribution of particular types of executive dysfunction, such as working memory or disinhibition, to observable behavior problems that hamper learning and achievement.

Overall, the present findings have important clinical implications for the evaluation of ADHD. Given the prevalence of the disorder, the negative psychosocial consequences associated with ADHD in adolescence and adulthood, the availability of effective treatments, and the possibility that some individuals may seek evaluation to obtain stimulant medication for illegal distribution, the importance of accurate clinical assessment has never been greater. The average 9-point deficit in overall cognitive ability observed in individuals with ADHD relative to control individuals complicates the interpretation of neuropsychological strengths and weaknesses. In particular, assessment approaches in which measures of specific functions are compared with overall ability may be insensitive, because overall ability is also decreased.[2] Instead, approaches comparing intact or less impaired functions to those thought to show greater impairment may provide more accurate detection. Although the present results suggest that there is not substantial differentiation of spared versus affected functions, aggregation across measures may enhance prediction. Future studies are needed that examine the utility of various methods of test battery interpretation in making an ADHD diagnosis.

The differential magnitude of effect sizes for intellectual and neuropsychological measures also has important implications for test selection. Specifically, full neuropsychological assessments may be excessive in circumstances in which only an accurate differential diagnosis is required. In these cases, a focused assessment using a few highly sensitive measures may be more practical and cost-effective than a full, detailed neuropsychological assessment. Similarly, it may be unnecessary to administer a large intellectual assessment battery to obtain an omnibus estimate of overall ability if specific neuropsychological tests are shown to possess greater discriminating power. In situations in which a full neuropsychological assessment is desired, clinicians should differentially weight measures when making diagnostic decisions. Furthermore, the large variability observed in mean effect sizes across neuropsychological measures highlights the need for a standardized approach to ADHD assessment. Most current assessment procedures have unknown predictive validity. For clinical practice to improve, studies are needed that directly assess the relative sensitivity and incremental validity of specific and global abilities. This work will substantially reduce evaluation time through the elimination of measures that are of limited importance to the referral question. It will also improve the diagnostic accuracy of neuropsychological assessments by providing a standard set of cognitive measures with known predictive ability.

[2] Ipsative interpretive approaches that compare specific subtest scores with the mean of an individual's performance across subtests are likely to be problematic in these situations (Glutting, Watkins, & Youngstrom, 2003).

References

Abikoff, H., Courtney, M. E., Szeibel, P. J., & Koplewicz, H. S. (1996). The effects of auditory stimulation on the arithmetic performance of children with ADHD and nondisabled children. *Journal of Learning Disabilities, 29,* 238–246.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

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

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

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

Barkley, R. A. (1997b). Behavioral inhibition, sustained attention, and executive functions: Constructing a unified theory of ADHD. *Psychological Bulletin, 121,* 65–94.

Barkley, R. A. (1998). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (2nd ed.). New York: Guilford Press.

Barkley, R. A., DuPaul, G. J., & McMurray, M. B. (1990). Comprehensive evaluation of attention deficit disorder with and without hyperactivity as defined by research criteria. *Journal of Consulting and Clinical Psychology, 58,* 775–789.

Barkley, R. A., Murphy, K., & Bush, T. (2001). Time perception and reproduction in young adults with attention-deficit/hyperactivity disorder. *Neuropsychology, 15,* 351–360.

Benton, A. L., & Hamsher, K. D. (1976). *Multilingual aphasia examination.* Iowa City: University of Iowa.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts.* New York: Guilford Press.

Carlson, C. L., Mann, M., & Alexander, D. K. (2000). Effects of reward and response cost on the performance and motivation of children with ADHD. *Cognitive Therapy and Research, 24,* 87–98.

Carlson, C. L., & Tamm, L. (2000). Responsiveness of children with attention deficit-hyperactivity/disorder to reward and response cost: Differential impact on performance and motivation. *Journal of Consulting and Clinical Psychology, 68,* 73–83.

Chhabildas, N., Pennington, B. F., & Willcutt, E. G. (2001). A comparison of neuropsychological profiles of the DSM–IV subtypes of ADHD. *Journal of Abnormal Child Psychology, 29,* 529–540.

Conners, C. K. (1994). *CPT: Conners Continuous Performance Test.* North Tonawanda, NY: Multi-Health Systems.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised.* Circle Pines, MN: American Guidance Service.

Dykman, R., & Ackerman, P. T. (1991). ADD and specific reading disability: Separate but often overlapping disorders. *Journal of Learning Disabilities, 24,* 96–103.

Faraone, S. V., Biederman, J., Spencer, T., Wilens, T., Seidman, L. J., Mick, E., et al. (2000). Attention-deficit/hyperactivity disorder in adults: An overview. *Biological Psychiatry, 48,* 9–20.

Fischer, M., Barkley, R. A., Edelbrock, C. S., & Smallish, L. (1990). The adolescent outcome of hyperactive children diagnosed by research criteria: II. Academic, attentional, and neuropsychological status. *Journal of Consulting and Clinical Psychology, 58,* 580–588.

Garcia-Sanchez, C., Estevez Gonzalez, A., Suarez Romero, E., & Junque, C. (1997). Right hemisphere dysfunction in subjects with attention-deficit disorder with and without hyperactivity. *Journal of Child Neurology, 12,* 107–115.

Glutting, J. J., Watkins, M., & Youngstrom, E. A. (2003). Multifactored and cross-battery assessments: Are they worth the effort? In C. R. Reynolds & R. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (2nd ed.). New York: Guilford Press.

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.

Goodyear, P., & Hynd, G. (1992). Attention deficit disorder with (ADD–H) and without (ADD–W) hyperactivity: Behavioral and neuropsychological differentiation. *Journal of Child Clinical Psychology, 21,* 273–305.

Gordon, M. (1983). *The Gordon Diagnostic System.* DeWitt, NY: Gordon Systems.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded.* Odessa, FL: Psychological Assessment Resources.

Jastak, S., & Wilkinson, G. S. (1984). *Wide Range Achievement Test—Revised.* Wilmington, DE: Jastak Associates.

Kagan, J. (1966). Reflection-impulsivity: The generality and dynamics of conceptual tempo. *Journal of Abnormal Psychology, 71,* 17–24.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence.* Boston, MA: Allyn & Bacon.

Kaufman, A. S. (1994). *Intelligent testing with the WISC–III.* New York: Wiley.

Kaufman, A. S., & Kaufman, N. L. (1990). *Manual for the Kaufman Brief Intelligence Test (K–BIT).* Circle Pines, MN: American Guidance Service.

Kemner, C., Verbaten, M. N., Cuperus, J. M., Camfferman, G., & van Engeland, H. (1995). Auditory event-related brain potentials in autistic children and three different control groups. *Biological Psychiatry, 38,* 150–165.

Kim, J., & Meuller, C. W. (1978). *Factor analysis: Statistical methods and practical issues* (Vol. 14). Newbury Park, CA: Sage.

Klorman, R., Hazel-Fernandez, L. A., Shaywitz, S. E., Fletcher, J. M., Marchione, K. E., Holahan, J. M., et al. (1999). Executive functioning deficits in attention-deficit/hyperactivity disorder are independent of oppositional defiant or reading disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 1148–1155.

Lezak, M. D. (1995). *Neuropsychological assessment.* New York: Oxford University Press.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

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,* 971–987.

Mariani, M. A., & Barkley, R. A. (1997). Neuropsychological and academic functioning in preschool boys with attention deficit hyperactivity disorder. *Developmental Neuropsychology, 13,* 111–129.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110,* 40–48.

Mitrushina, M. M., Boone, K. B., & D'Elia, L. F. (1999). *Handbook of normative data for neuropsychological assessment.* New York: Oxford Press.

Murphy, K. R., Barkley, R. A., & Bush, T. (2001). Executive functioning and olfactory identification in young adults with attention-deficit/hyperactivity disorder. *Neuropsychology, 15,* 211–220.

Oosterlaan, J., Logan, G. D., & Sergeant, J. A. (1998). Response inhibition in AD/HD, CD, comorbid AD/HD + CD, anxious, and control children: A meta-analysis of studies with the stop task. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39,* 411–425.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157–159.

Ozonoff, S., & Jensen, J. (1999). Specific executive function profiles in three neurodevelopmental disorders. *Journal of Autism and Developmental Disorders, 29,* 171–177.

Quay, H. C. (1997). Inhibition and attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology, 25,* 7–13.

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 Child Clinical Psychology, 29,* 555–568.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8,* 271–276.

Rey, A., & Osterrieth, P. A. (1993). Translations of excerpts from Andre Rey's psychological examination of traumatic encephalopathy and P. A. Osterrieth's The Complex Figure Copy Test. *Clinical Neuropsychologist, 7,* 4–21.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Vol. 6). Newbury Park, CA: Sage.

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99,* 400–406.

Rucklidge, J. J., & Tannock, R. (2001). Psychiatric, psychosocial, and cognitive functioning of female adolescents with ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 530–540.

Sattler, J. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.

Seidman, L. J., Biederman, J., Monuteaux, M. C., Doyle, A. E., & Faraone, S. V. (2001). Learning disabilities and executive dysfunction in boys with attention-deficit/hyperactivity disorder. *Neuropsychology, 15,* 544–556.

Semrud-Clikeman, M., Biederman, J., Sprich-Buckminster, S., Lehman, B. K., Faraone, S. V., & Norman, D. (1992). Comorbidity between ADDH and learning disability: A review and report in a clinically referred sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 31,* 439–448.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9,* 371–377.

Slomka, G. (1998). Attention-deficit/hyperactivity disorder. In P. J. Snyder & P. D. Nussbaum (Eds.), *Clinical neuropsychology: A pocket handbook for assessment.* Washington, DC: American Psychological Association.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary.* New York: Oxford University Press.

Stroop, J. P. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Technical manual, Stanford–Binet Intelligence Scale* (4th ed.). Chicago, IL: Riverside.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale—Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence—Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1999). *Manual for the Wechsler Abbreviated Scale of Intelligence.* San Antonio, TX: Psychological Corporation.

Wilkinson, G. S. (1993). *The Wide Range Achievement Test—Third Edition.* San Antonio, TX: Psychological Corporation.

Woods, S. P., Lovejoy, D. W., & Ball, J. D. (2002). Neuropsychological characteristics of adults with ADHD: A comprehensive review of initial studies. *The Clinical Neuropsychologist, 16,* 12–34.

Youngstrom, E. A., Findling, R. L., & Calabrese, J. R. (2003). Who are the comorbid adolescents? Agreement between psychiatric diagnosis, parent, teacher, and youth report. *Journal of Abnormal Child Psychology, 31,* 231–245.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.

Received February 26, 2003
Revision received September 9, 2003
Accepted September 16, 2003 ■