# CONSTRUCT VALIDITY OF THE KAUFMAN BRIEF INTELLIGENCE TEST, WECHSLER INTELLIGENCE SCALE FOR CHILDREN-THIRD EDITION, AND ADJUSTMENT SCALES FOR CHILDREN AND ADOLESCENTS

Gary L. Canivez Eastern Illinois University

> Ryan Neitzel School District 87 Bloomington, Illinois

Blake E. Martin Mid-State Special Education Taylorville, Illinois

The present study reports data supporting the construct validity of the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990), the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991), and the Adjustment Scales for Children and Adolescents (ASCA; McDermott, Marston, & Stott, 1993) through convergent and discriminant comparisons in a sample of 207 students receiving special education evaluations. Results were as hypothesized, with high and statistically significant correlations between the K-BIT and WISC-III, supporting convergent validity. Moderate and statistically significant correlations were obtained between the two intelligence measures (K-BIT and WISC-III) and measures of academic achievement (WIAT, WIAT-II, WJ-R ACH, or WJ-3 ACH) at levels typical of ability-achievement correlations. Correlations between the two intelligence measures (K-BIT and WISC-III) and the ASCA, a measure of child psychopathology, were low to near zero, supporting discriminant validity. Further discriminant evidence of construct validity was provided by the low to near zero correlations between the ASCA and the measures of academic achievement.

Investigations of construct validity of psychological assessment instruments are of critical importance for professional psychologists in order to determine interpretability of psychological tests. In fact, the *Standards for Educational and Psychological Testing* (AERA, APA, NCME; 1999) require that test interpretation methods demonstrate empirical support and such empirical support comes

The authors thank Maria Rayborn for assistance in data collection and Tara Plopper, David Long, Kim Victor, Erin Willenborg, Amanda Kearney, Rachel Witte, Susan Kapper, Andrew Tourville, and Timothy Allen for assistance in scoring ASCA rating forms and computer data entry.

Preliminary versions of this study were presented at the 2003 Annual Convention of the National Association of School Psychologists and the 2003 Annual Convention of the American Psychological Association.

Correspondence concerning this article and requests for tables should be addressed to Gary L. Canivez, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920. Information on the ASCA may be obtained at <a href="http://www.uxl.eiu.edu/~cfglc/ASCA.html">http://www.uxl.eiu.edu/~cfglc/ASCA.html</a>.

from studies of construct validity. Many investigations in the published literature might be considered or classified as convergent validity, where tests that measure the same construct are compared and expected to demonstrate high and statistically significant correlations and nonsignificant mean differences between similar scales. Campbell and Fiske (1959) argued that it was also important to compare tests designed to measure *different* constructs and obtain low to near zero correlations for the establishment of construct validity of psychological measures. They termed this type of validity *discriminant* validity. The present study investigated the convergent and discriminant validity of some commonly used tests in psychological assessment of children and adolescents.

The Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) was one of the most widely used measures of intelligence (Stinnett, Havey, & Oehler-Stinnett, 1994), and it was an important component of the special education eligibility determination process for millions of students (Gresham & Witt, 1997; Kamphaus, Petoskey, & Rowe, 2000). As a comprehensive measure of intellectual abilities, the WISC-III required approximately 1 hour to 1½ hours to administer; however, there are several situations (i.e., screening, research, reevaluation) where such comprehensive assessment may not be necessary (Canivez, 1995, 1996a, 1996b; Kaufman & Kaufman, 1990). Numerous short forms of the Wechsler scales have been developed over the years (Sattler, 1992, 2001; Silverstein, 1990) to reduce the assessment time when using comprehensive intelligence tests like the Wechsler scales.

Several statistical and utilization problems are associated with short forms of comprehensive intelligence tests. Silverstein (1990) argued that short form correlations with scores such as VIQ, PIQ, FSIQ, VCI, and POI would be spuriously high due to their inclusion in calculating the IQ or Index score. Another problem with short forms is that they are developed utilizing standardization data where the individuals are administered the entire test in a standardized subtest order and the resulting scores may not correspond if only the short form subtests were administered in isolation. Kaufman and Kaufman (2001) argued that the development and standardization of well-normed, reliable, and valid brief tests of intelligence like the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990) and the Wechsler Abbreviated Scale of Intelligence (WASI; The Psychological Corporation, 1999) eliminated the need to create short forms of more comprehensive intelligence tests such as the WISC-III.

Numerous validity studies have generally supported the concurrent and convergent validity of the WISC-III (Bell, Rucker, Finch, & Alexander, 2002; Caravajal, Hayes, Lackey, & Rathke, 1993; DiCerbo & Barona, 2000; Dumont, Cruse, Price, & Whelley, 1996; Law & Faison, 1996; Lukens & Hurrell, 1996; Prewett & Matavich, 1994; Rust & Lindstrom, 1996; Rust & Yates, 1997; Saklofske, Schwean, Yaculic, & Quinn, 1994; Vo, Weisenberger, Becker, & Jacob-Timm, 1999; Wechsler, 1991) and the factorial validity of the WISC-III (Allen & Thorndike, 1995; Grice, Krohn, & Logerquist, 1999; Keith & Witta, 1997; Konold, Kush, & Canivez, 1997; Kush, 1996; Kush & Watkins, 1997; Kush, Watkins, Ward, Canivez, & Worrell, 2001; Logerquist-Hansen & Barona, 1994; Reynolds & Ford, 1994; Roid, Prifitera, & Weiss, 1993; Roid & Worrall,

1997; Sattler, 2001; Scardapane, 1996; Thorndike, 1992; Watkins, Greenawalt, & Marcell, 2002; Wechsler, 1991). Additional studies have provided support for the concurrent and predictive validity of the K-BIT (Axelrod & Naugle, 1988; Bowers & Pantle, 1998; Canivez, 1995, 1996a, 1996b; Chin et al., 2001; Donders, 1995; Eisenstein & Engelhart, 1997; Grados & Russo-Garcia, 1999; Hayes, 1999; Hays, Reas, & Shaw, 2002; Kaufman & Kaufman, 1990; Levinson & Folino, 1994; Naugle, Chelune, & Tucker, 1993; Prewett, 1992a, 1992b, 1995; Prewett & McCaffery, 1993; Slate, Graham, & Bower, 1996; Thompson, Browne, Schmidt, & Boer, 1997; Webber & McGillivray, 1998). Finally, concurrent, convergent, and divergent (discriminant) validity (Canivez & Bordenkircher, 2002; Canivez & Rains, 2002; McDermott, 1993, 1994, 1995) and factorial validity (Canivez, 2004; McDermott, 1993, 1994) studies have supported the Adjustment Scales for Children and Adolescents (ASCA; McDermott, Marston, & Stott, 1993).

Reviews of research on the WISC-III, K-BIT, and ASCA have revealed a variety of convergent, concurrent, and factorial validity studies, but none has simultaneously examined both convergent and discriminant validity of all three. The purpose of the present study was to go beyond the typical convergent validity or concurrent validity studies by examining discriminant validity as well. The present study describes an investigation of construct validity of the K-BIT, WISC-III, and ASCA through simultaneous comparisons in a sample of students referred for special education evaluations. Convergent evidence of construct validity would be supported by statistically significant and high correlations between the K-BIT and WISC-III in addition to no statistically significant or meaningful mean differences between the global IQ scores. Discriminant evidence of construct validity would be supported by low to near zero correlations between the K-BIT and ASCA and between the WISC-III and ASCA given that these instruments measure quite different and theoretically independent constructs (intelligence vs. psychopathology). Because data were obtained from comprehensive psychoeducational evaluations, individually administered academic achievement test data were also available and served as an additional source of comparison for the intelligence and psychopathology measures. It was hypothesized that the correlations between the intelligence (K-BIT and WISC-III) measures and the academic achievement measures would be moderately high (but not as high as the K-BIT-WISC-III correlations), statistically significant, and near the median intelligence-achievement correlations reported by Heath and Kush (1991) and Naglieri and Bornstein (2003). Further evidence of discriminant validity would be shown by low, near zero, and statistically nonsignificant correlations between the achievement measures and the ASCA.

# METHOD

#### Participants

A total of 207 public school children and adolescents who were evaluated by multidisciplinary evaluation teams for special education consideration were participants in the present study. Parents or legal guardians provided informed consent for the evaluations consistent with The Individuals with Disabilities Education Act (IDEA; Public Law 105-17), and all data were collected anonymously to protect the identity of individual students. Of the 207, 130 (62.8%) were male, 77 (37.2%) were female, 167 (80.7%) were Caucasian, 33 (15.9%) were African American, and 7 (3.4%) were Hispanic/Latino. Students ranged in grade from kindergarten through grade 10<sup>1</sup> and in age from 6 to 16.83 years (M = 10.51, SD = 2.55). The sample was dominated by students with disabilities; however, 43 (20.8%) students were not classified by multidisciplinary evaluation teams as having a disability. As is typically found in special education, students with specific learning disability (n = 118, 57%) were the most prevalent disability group. Students with other disabilities included 23 (11.1%) with mental retardation, 15 (7.2%) with serious emotional disability, 6 (2.9%) with attention deficit hyperactivity disorder, 1 (0.5%) with autism, and 1 (0.5%) with serious emotional disability and specific learning disability.

#### Instruments

Kaufman Brief Intelligence Test. The Kaufman Brief Intelligence Test (K-BIT) is a "brief, individually administered measure of the verbal and nonverbal intelligence of a wide range of children, adolescents, and adults, spanning the ages of 4 to 90 years" (Kaufman & Kaufman, 1990, p. 1). It is comprised of two subtests, Vocabulary (Expressive Vocabulary and Definitions) and Matrices, and takes approximately 15 to 30 minutes to administer. The K-BIT was standardized on a representative sample (N=2,022) closely approximating 1990 United States Census data on variables of gender, geographic region, socioeconomic status, and race/ethnic group. By including both verbal and nonverbal subtests (most previous brief intelligence tests focused on one or the other), the K-BIT is able to measure two different skill areas, which purportedly allows the examiner to assess verbal-nonverbal discrepancies as is done with the Wechsler scales (Kaufman & Kaufman, 1990). Kaufman and Kaufman were cautious about the interpretation of significant ( $\alpha = .05$  or .01) Vocabulary–Matrices discrepancies and recommended that practitioners first evaluate the size of the discrepancy based upon a selected "abnormal amount of scatter" observed in the standardization sample; and second, "not attempt to interpret the clinical, psychoeducational, or neuropsychological implications" of such differences (p. 46). Kaufman and Kaufman recommended that such differences be used to formulate hypotheses about the individual's unique pattern of abilities that should be further investigated with a comprehensive assessment. Canivez (1995), however, presented data questioning the utility of this comparison.

Split-half internal consistency reliability estimates across the entire age range for the K-BIT IQ Composite, Vocabulary, and Matrices scores were high, ranging from .88 to .98 (mean r = .94), .89 to .98 (mean r = .93), and .74 to .95 (mean r = .88), respectively. Test-retest stability estimates for the IQ Composite, Vocabulary, and Matrices scores with four age samples ranged from .92 to .95 (mean r = .94), .86 to .97 (mean r = .94), and .80 to .92 (mean r = .85), respectively (Kaufman & Kaufman, 1990).

<sup>&</sup>lt;sup>1</sup>A table presenting distributions of all demographic characteristics is available upon request.

Table 1

Correlations between Comparable Indexes of the K-BIT, WISC-R, WAIS-R, WISC-III, SBIS-FE, WASI, and SILS

		IQ Index			
Scales/Studies	N	Verbal	Nonverbal	Overall	
K-BIT—WAIS-R				der feit die einer einfen som som	
Axelrod & Naugle (1998)	200	.83	.77	.88	
Eisenstein & Engelhart (1997)	64			.73ª	
Kaufman & Kaufman (1990)	64	.60	.52	.75	
Naugle, Chelune, & Tucker (1993)	200	.83	.77	.88	
K-BIT-WISC-R					
Kaufman & Kaufman (1990)	35	.78	.50	.80	
Prewett (1992a)	35	.83	.70	.81	
Prewett (1992b)	40	.70	.29	.64	
Slate, Graham, & Bower (1996)	44	.79	.74	.80	
Webber & McGillivray (1998)	107	.63	.53	.73	
K-BIT-WISC-III					
Canivez (1995)	137	.80	.74	.87	
Canivez (1996a, 1996b)	75	.72	.64	.82	
Chin, Ledesma, Cirino, Sevcik, Morris,					
Frijters, & Lovett (2001)	65	.60	.48	.63	
Donders (1995)	47	.79	.52	.74	
Grados & Russo-Garcia (1999)	35	.65	.80	.86	
Levinson & Folino (1994)	29	.35	.35	.53	
Prewett (1995)	50			.78	
K-BIT-WISC-III Short Form					
Thompson, Browne, Schmidt, & Boer (1997)	42	.79	.45	.74	
K_RIT_SRIS_FF					
Prewett & McCaffery (1993)	75		•	.81	
K-BII-MAI-SF	126			69	
Hayes (1999)	126			.00	
Prewell (1995)	50			./0	
K-BIT-CPM					
Prewett & McCaffery (1993)	75			.81	
K-BIT-WASI					
Webber & McGillivray (1998)	107		.53	.52	
Hays, Reas, & Shaw (2002)	85			.89	
K-BIT-SILS					
Bowers & Pantle (1998)	30	.83	.20	.77	
Bowers & Pantle (1998)	50	.77	.77	.83	

Note.—K-BIT = Kaufman Brief Intelligence Test, WAIS-R = Wechsler Adult Intelligence Scale-Revised, WISC-R = Wechsler Intelligence Scale for Children-Revised, WISC-III = Wechsler Intelligence Scale for Children-Third Edition, SB-FE = Stanford Binet Intelligence Scale-Fourth Edition, MAT-SF = Matrix Analogies Test-Short Form, CPM = Coloured Progressive Matrices, WASI = Wechsler Abbreviated Scale of Intelligence, SILS = Shipley Institute for Living Scales. <sup>a</sup> intraclass r.

Concurrent validity studies reported in the K-BIT manual (Kaufman & Kaufman, 1990) supported the K-BIT with comparisons to the Test of Nonverbal Intelligence (TONI; Brown, Sherbenou, & Johnsen, 1990), Slosson Intelligence Test (SIT; Jensen & Armstrong, 1985), Kaufman Assessment

Battery for Children (K-ABC; Kaufman & Kaufman, 1983), Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974), and Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981). Independent validity research on the K-BIT has also provided empirical support. Table 1 presents a summary of correlation coefficients from K-BIT concurrent or convergent validity comparisons with comprehensive intelligence tests such as the WISC-R (Wechsler, 1974) (Prewett, 1992a, 1992b; Slate, Graham, & Bower, 1996; Webber & McGillivray), the WAIS-R (Wechsler, 1981) (Axelrod & Naugle, 1988; Eisenstein & Engelhart, 1997; Naugle, Chelune, & Tucker, 1993), the Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE; Thorndike, Hagen, & Sattler, 1986) (Prewett & McCaffery, 1993), the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) (Canivez, 1995, 1996a, 1996b; Chin et al., 2001; Donders, 1995; Grados & Russo-Garcia, 1999; Levinson & Folino, 1994; Prewett, 1995), and with other brief intelligence measures such as the WISC-III short form (Picture Completion, Information, Block Design, and Vocabulary; Sattler, 1992, 2001) (Thompson, Browne, Schmidt, & Boer, 1997), the Shipley Institute for Living Scale (SILS; Zachary, 1986) (Bowers & Pantle, 1998), the Wechsler Abbreviated Scale of Intelligence (WASI; The Psychological Corporation, 1999) (Hays, Reas, & Shaw, 2002; Webber & McGillivray, 1998), the Matrix Analogies Test-Short Form (MAT-SF; Naglieri, 1985) (Hayes, 1999; Prewett, 1995), and Raven's Coloured Progressive Matrices (CPM; Raven, 1956) (Prewett & McCaffery, 1993).

Wechsler Intelligence Scale for Children-Third Edition. The Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) is an individually administered test of intellectual abilities for children aged 6 years through 16 years 11 months. The WISC-III yields three composite IQs, viz., Verbal (VIQ), Performance (PIQ), and Full Scale (FSIQ), and four optional factor-based index scores, viz., Verbal Comprehension (VCI), Perceptual Organization (POI), Freedom from Distractibility (FDI), and Processing Speed (PSI). The WISC-III was standardized on a representative sample (N = 2,200) closely approximating the 1988 United States Census on gender, parent education (SES), race/ethnicity, and geographic region. Internal consistency reliability estimates for the three IQ and four Index scores were high, ranging from .80 to .97 within the 11 age levels with 55 of 77 (71%) coefficients = .90. Average test-retest stability estimates for the three IQ and four Index scores were also high, ranging from .82 to .94. Long-term (3-year) stability of the WISC-III IQ scores and the four-factor model has also been supported (Canivez & Watkins, 1998, 1999, 2001; Watkins & Canivez, 2001).

Concurrent validity studies generally found moderately high correlations with other intellectual ability measures, and VIQ tended to correlate higher with verbal ability measures than nonverbal ability measures, whereas PIQ tended to correlate higher with nonverbal ability measures than verbal ability measures (Wechsler, 1991), as expected. Exploratory and confirmatory factor analyses of the WISC-III standardization sample (Wechsler, 1991) suggested a four-factor model, which was replicated with the WISC-III Canadian normative sample (Roid & Worrall, 1997). A similar four-factor structure was given qualified support by Keith and Witta (1997), but others (Allen & Thorndike, 1995; Sattler, 2001; Thorndike, 1992) have suggested that a three-factor model best describes the WISC-III. When Symbol Search was excluded from analyses, a three-factor model was observed across different ages and factor analytic techniques (Reynolds & Ford, 1994). Independent samples have also produced some contradictory results, ranging from four to two factors (Grice, Krohn, & Logerquist, 1999; Konold, Kush, & Canivez, 1997; Kush, 1996; Kush & Watkins, 1997; Kush et al., 2001; Logerquist-Hansen & Barona, 1994; Roid et al., 1993; Scardapane, 1996; Watkins et al., 2002). Generally speaking, strong support for the Verbal and Performance dimensions has been replicated across all studies but controversy and divergent results still exist with the two smaller factors (Freedom from Distractibility and Processing Speed).

In addition to factorial validity, several concurrent and convergent validity studies have supported the WISC-III with comparisons to other measures of intelligence like the Differential Abilities Scale (DAS; Elliott, 1990) (DiCerbo & Barona, 2000; Dumont et al., 1996), the SB:FE (Thorndike et al., 1986) (Caravajal et al., 1993; Lukens & Hurrell, 1996; Prewett & Matavich, 1994; Rust & Lindstrom, 1996; Saklofske et al., 1994), the Slosson Full Range Intelligence Test (S-FRIT; Algozzine, Eaves, Mann, & Vance, 1993) (Bell et al., 2002), the Kaufman Adolescent and Adult Intelligence Test (KAIT; Kaufman & Kaufman, 1993) (Law & Faison, 1996; Vo et al, 1999), and the K-ABC (Kaufman & Kaufman, 1983) (Rust & Yates, 1997). Schultz (1997) found that the WISC-III FSIQ produced correlations with the Woodcock-Johnson Revised Tests of Achievement (WJ-R ACH; Woodcock & Mather, 1989) at levels frequently observed between intelligence and achievement tests (Heath & Kush, 1991; Naglieri & Bornstein, 2003).

Adjustment Scales for Children and Adolescents. The Adjustment Scales for Children and Adolescents (ASCA; McDermott, Marston, & Stott, 1993) is an objective behavior rating instrument completed by a student's classroom teacher and designed for use with all noninstitutionalized youths aged 5 through 17 (grades K through 12). The ASCA consists of 156 behavioral descriptions within 29 specific situations where teachers may observe students' behaviors. Of the 156 items, 97 are scorable for psychopathology and, based on factor analyses, singularly assigned to one of six core syndromes (Attention-Deficit/Hyperactive, Solitary Aggressive-Provocative, Solitary Aggressive-Impulsive, Oppositional Defiant, Diffident, and Avoidant) or two supplementary syndromes (Delinquent and Lethargic/Hypoactive). The core syndromes are combined to form two composite indexes: Overactivity (Attention-Deficit Hyperactive, Solitary Aggressive-Provocative, Solitary Aggressive-Impulsive, and Oppositional Defiant syndromes) and Underactivity (Diffident and Avoidant syndromes). Core syndromes, supplementary syndromes, and overall adjustment scales are reported as normalized T scores (M = 50, SD = 10) and percentiles based on a nationally representative standardization sample of 1,400 youths, blocked according to gender, age, and grade level and stratified proportionately according to national region, community size, race/ethnicity, parent education, family structure, and handicapping condition. In general, psychometric properties of the ASCA are acceptable and meet standards for both group and individual decision making (Canivez, 2001; Salvia & Ysseldyke, 1995).

Extensive evidence for ASCA score reliability and validity is presented in the ASCA manual (McDermott, 1994) and independent studies. Internal consistency estimates for the total standardization sample ranged from .68 to .86 for the six core syndromes and two supplementary syndromes and equaled .92 for the Overactivity scale and .82 for the Underactivity scale. Test-retest stability coefficients over a 30-school day interval ranged from .66 to .91 for the six core syndromes and from .75 to .79 for the Overactivity and Underactivity scales. No significant differences in mean T scores were observed across the retest interval. Canivez, Perry, and Weller (2001) also found statistically significant stability coefficients for the ASCA overall adjustment scales, core syndromes, and supplemental syndromes over a 90-day retest interval, and mean changes were less than 0.8 raw score points. Canivez et al. (2001) also found statistically significant stability for the ASCA syndromic profiles and discriminant classifications, two additional methods of score interpretation. McDermott (1994) and Watkins and Canivez (1997) reported statistically significant interrater agreement for ASCA syndrome T scores. Statistically significant correlations were found for the core syndromes and global adjustment scales, and no statistically or clinically significant mean differences were found between raters. Canivez and Watkins (2002) reported statistically significant interrater agreement for ASCA Syndromic Profile Classifications, and Canivez, Watkins, and Schaefer (2002) reported statistically significant interrater agreement for ASCA Discriminant Classifications.

Evidence of convergent, divergent (discriminant), discriminative, and factorial validity of the ASCA has also been reported. McDermott (1993, 1994) reported statistically significant convergent validity coefficients between the ASCA and the Revised Conners Teacher Rating Scale (CTRS; Trites, Blouin, & Laprade, 1982). The low and near zero correlations between the Overactive and Underactive core syndromes of the ASCA supported the divergent (discriminant) validity and independence of these two dimensions (McDermott, 1993; 1994). Correlations among similar psychological dimensions or constructs between the ASCA and Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983) were also statistically significant and moderately high (McDermott, 1993, 1994). Canivez and Bordenkircher (2002) and Canivez and Rains (2002) found convergent and divergent (discriminant) validity support in comparing the ASCA and the Preschool and Kindergarten Behavior Scales (PKBS; Merrell, 1994a) among randomly selected preschool, kindergarten, and first-grade children. Divergent (discriminant) validity was observed, with low to near zero correlations between the ASCA Overactivity global adjustment syndrome and related core syndrome scores and the PKBS Internalizing composite and subscale scores. Additional divergent (discriminant) evidence of construct validity for the ASCA was reported by McDermott (1995), who found low negative correlations (except one comparison) between the ASCA and the Differential Abilities Scale (DAS; Elliott, 1990). Psychological adjustment as measured by the ASCA accounted for no more than 6% of the variability in

ability or achievement as measured by the DAS. McDermott (1994) and McDermott et al. (1995) showed that the ASCA core syndromes also demonstrated good diagnostic accuracy (approximately 80% correct classification) in differentiating students with emotional disturbance from age-, gender-, race-, and grade level-matched normal students, as well as separate groups of learning-disabled, speech/language-disabled, and gifted students. Positive predictive power estimates also exceeded a recommended standard (.75) for diagnostic tests (Landau, Milich, & Widiger, 1991). Canivez and Sprouls (in press) found the ASCA to differentiate students meeting independent ADHD criteria from random and matched normal students with positive predictive power of .94. Other diagnostic efficiency statistics (Kessel & Zimmerman, 1993) were also highly supportive.

Exploratory factor analyses and confirmatory analyses reported by McDermott (1993, 1994) indicated that the ASCA items are best explained by an eight-factor model with six factors (core syndromes) generalizing across gender, race/ethnicity, and age, whereas two factors (supplemental syndromes) were appropriate for specific subgroups in the population. Factor analyses of the six core syndromes produced a two-factor solution (Overactivity and Underactivity), which appears similar to the two-dimensional model (conduct problem/externalizing vs. withdrawal/internalizing) of child psychopathology frequently obtained in the assessment literature (Achenbach, 1991; Achenbach & Edelbrock, 1983; Cicchetti & Toth, 1991; Merrell, 1994a, 1994b, 2003; Quay, 1986; Reynolds & Kamphaus, 1992, 2004). Core syndrome specificity estimates were also shown to be higher than error estimates and indicated that the separate core syndromes can be meaningfully interpreted (McDermott, 1993, 1994). Canivez (2004) also found that the six ASCA core syndromes produced the same two-factor solution (Overactivity and Underactivity) in an independent sample of 1,020 children and adolescents, and internal consistency estimates were similar to those obtained in the standardization data.

#### Procedure

Students were administered the K-BIT and WISC-III in a random counterbalanced order, during the same test session, as part of a comprehensive psychoeducational evaluation to determine disability in initial special education evaluations or reevaluations. Evaluations were conducted by four state and nationally certified school psychologists. Students for whom the school psychologists selected the WISC-III as the intelligence test to be administered in their evaluation of cognitive skills were selected for participation in the present study and were also administered the K-BIT. Individually administered standardized achievement tests-Woodcock-Johnson Revised Tests of Achievement (WJ-R ACH; Woodcock & Mather, 1989); Woodcock-Johnson III Tests of Achievement (WJ-III ACH; Woodcock, McGrew, & Mather, 2001); Wechsler Individual Achievement Test (WIAT; The Psychological Corporation, 1992); or Wechsler Individual Achievement Test-II (WIAT-II; The Psychological Corporation, 2001)—were also administered as part of the comprehensive psychoeducational evaluations. The ASCA was completed by the student's regular classroom teacher or, in the case of students already in special education classes, the teacher who spent the most time with the child. For the older students, school psychologists had content-area teachers provide ASCA ratings rather than "home room" or "home base" teachers. School psychologists scored all tests, and data were coded anonymously and sent to the first author. ASCA rating forms were also sent to the first author, who had trained research assistants score and enter all data for analyses.

# Data Analyses

Pearson product-moment correlation coefficients were calculated between the K-BIT Vocabulary, Matrices, and IQ Composite standard scores and the WISC-III VIQ, PIQ, FSIQ, VCI, POI, FDI, and PSI scores. The correlation between the K-BIT Vocabulary–Matrices discrepancy and the WISC-III VIQ–PIQ discrepancy was also calculated because these comparisons are frequently conducted and interpreted (Kaufman, 1994; Sattler, 2001; Wechsler, 1991). Dependent t tests for differences between means were also used to examine *level* differences (McDermott, 1988) between similar subtests and composite scores of the K-BIT and WISC-III. Pearson product-moment correlations were also calculated between the K-BIT and WISC-III and the achievement tests, between the K-BIT and WISC-III and the ASCA, and between the ASCA and the achievement tests.

# RESULTS

#### Intelligence Test Comparisons

Table 2 presents the descriptive statistics and correlations between the K-BIT and WISC-III. Sample sizes differed because the school psychologists did not universally administer all 12 WISC-III subtests that allow for calculation of all IQ and Index scores. As seen in Table 2, correlations ranged from .40 to .89  $(Mdn_r = .67)$ . The correlation between the K-BIT IQ Composite and WISC-III FSIQ (r = .89) was statistically significant and high and indicated a large proportion of shared variance  $(r^2 = .79)$ . The K-BIT Vocabulary subtest was significantly correlated with the WISC-III VIQ (r = .85) and with the WISC-III VCI (r = .84). Statistically significant but somewhat lower correlations were observed between the K-BIT Matrices subtest and the WISC-III PIQ (r = .67) and WISC-III POI (r = .56). The correlation between the K-BIT Vocabulary–Matrices discrepancy and WISC-III VIQ–PIQ discrepancy (r = .38) was statistically significant but was low to moderate in magnitude and accounted for only 14% shared variance.

Results of the dependent t tests for differences between mean scores of similar K-BIT and WISC-III indexes found statistically significant mean differences between the K-BIT IQ Composite score (M = 89.39, SD = 15.60) and the WISC-III FSIQ (M = 87.19, SD = 14.87), t(205) = -4.43, p < .05, Glass's  $\Delta = .15$ ; and between the K-BIT Vocabulary subtest (M = 89.20, SD = 15.25) and the WISC-III VIQ (M = 86.85, SD = 14.68), t(205) = -4.11, p < .05, Glass's  $\Delta = .16$ . These mean differences of approximately 2 IQ points were well within the standard error of measurement of both measures and also represented small effect sizes (Cohen, 1992) based on Glass's  $\Delta$  (Glass & Hopkins, 1996). Thus these differences were not clinically significant. Statistically significant mean differences were not observed between the K-BIT Vocabulary and WISC-III VCI, the K-BIT Matrices and WISC-III PIQ, or the K-BIT Matrices and WISC-III POI.

						<u> </u>		
				Kaufman Brief Intelligence Test				
WISC-III	n	М	SD		. Vocabulary	Matrices	IQ Composite	
VIQ	206	86.85	14.68		.85	.57	.83	
PIQ	206	89.84	14.98		.72	.67	.82	
FSIQ	206	87.19	14.87		.85	.67	.89	
VCI	202	88.43	14.66		.84	.55	.82	
POI	202	89.77	15.92		.65	.56	.73	
FDI	124	84.10	12.64		.65	.43	.70	
PSI	97	89.90	15.65		.57	.40	.67	
				М	89.20	90.79	89.39	
				SD	15.25	16.67	15.60	

Table 2 Descriptive Statistics and Pearson Product-Moment Correlations between Intelligence Scales

Note.—WISC-III = Wechsler Intelligence Scale for Children-Third Edition, VIQ = Verbal IQ, PIQ = Performance IQ, FSIQ = Full Scale IQ, VCI = Verbal Comprehension Index, POI = Perceptual Organization Index, FDI = Freedom from Distractibility Index, PSI = Processing Speed Index. Sample sizes vary depending on what WISC-III subtests were reportedly administered.

All correlations significant (p < .05) after controlling for the family-wise error rate with Bonferroni correction, p < .002.

IQ Scores		ores			
WISC-III	BR	RC	MC	MR	WE
VIQ	.59ª	.65 <sup>f</sup>	.65°	.70 <sup>d</sup>	.53 <sup>j</sup>
PIQ	.52ª	.62 <sup>f</sup>	.56 <sup>c</sup>	.69 <sup>d</sup>	.48 <sup>j</sup>
FSIQ	.59ª	.68 <sup>f</sup>	.65 <sup>c</sup>	.74 <sup>d</sup>	.54 <sup>j</sup>
VCI	.56 <sup>e</sup>	.61 <sup>i</sup>	.58 <sup>g</sup>	.64 <sup>h</sup>	.52
POI	.47 <sup>e</sup>	.57 <sup>i</sup>	.48 <sup>g</sup>	.63 <sup>h</sup>	.44 <sup>m</sup>
FDI	.49 <sup>n</sup>	.49 <sup>p</sup>	.57°	.53°	.40 <sup>q</sup>
PSI	.42'	.49 <sup>t</sup>	.48 <sup>s</sup>	.46 <sup>s</sup>	.42"
K-BIT					
Vocabulary	.58 <sup>b</sup>	.61 <sup>g</sup>	.62 <sup>d</sup>	.68 <sup>e</sup>	.54 <sup>k</sup>
Matrices	.45 <sup>b</sup>	.46 <sup>g</sup>	.42 <sup>d</sup>	.55 <sup>e</sup>	.38 <sup>k</sup>
IQ Composite	.60 <sup>b</sup>	.66 <sup>g</sup>	.61 <sup>d</sup>	.72 <sup>e</sup>	.55 <sup>k</sup>
М	84.12	83.01	82.37	85.43	84.33
SD	12.62	13.28	14.20	12.94	13.14

 Table 3

 Pearson Product-Moment Correlations between Intelligence and Achievement Scales

Note.—BR = Basic Reading, RC = Reading Comprehension, MC = Mathematics Calculation, MR = Mathematics Reasoning, WE = Written Expression, WISC-III = Wechsler Intelligence Scale for Children-Third Edition, VIQ = Verbal IQ, PIQ = Performance IQ, FSIQ = Full Scale IQ, VCI = Verbal Comprehension Index, POI = Perceptual Organization Index, FDI = Freedom from Distractibility Index, PSI = Processing Speed Index.

 ${}^{a}n = 207$ .  ${}^{b}n = 206$ .  ${}^{c}n = 205$ .  ${}^{d}n = 204$ .  ${}^{e}n = 203$ .  ${}^{f}n = 202$ .  ${}^{g}n = 201$ .  ${}^{h}n = 200$ .  ${}^{i}n = 198$ .  ${}^{i}n = 171$ .  ${}^{k}n = 170$ .  ${}^{i}n = 168$ .  ${}^{m}n = 167$ .  ${}^{n}n = 124$ .  ${}^{o}n = 123$ .  ${}^{p}n = 119$ .  ${}^{q}n = 106$ .  ${}^{r}n = 97$ .  ${}^{s}n = 96$ .  ${}^{t}n = 93$ .  ${}^{u}n = 82$ . Samples vary depending on which WISC-III subtests and achievement subtests were reportedly administered.

All correlations significant (p < .05) after controlling for the family-wise error rate with Bonferroni correction within achievement areas, p < .005.

#### Intelligence and Achievement Test Comparisons

Table 3 presents the descriptive statistics and correlation coefficients between the WISC-III and K-BIT and the achievement test scores. Ability-achievement correlations for the WISC-III ranged from .40 to .74 ( $Mdn_r = .56$ ).

Ability-achievement correlations for the K-BIT ranged from .38 to .72 ( $Mdn_r = .58$ ). All correlations were statistically significant (p < .05) after controlling for the family-wise error rate with Bonferroni correction (Dunn, 1961) within the achievement areas. Further, all correlations were moderate in magnitude.

### Psychopathology and Intelligence Test Comparisons

As hypothesized, all correlations between the WISC-III and the ASCA were low to near zero and ranged from -.18 to .07  $(Mdn_r = -.10)^2$ . The highest relationship between the WISC-III and ASCA indicated only 3.2% shared variance. As hypothesized, all correlations between the K-BIT and the ASCA were also low to near zero and ranged from -.17 to .04  $(Mdn_r = -.10)$ . The highest relationship between the K-BIT and ASCA resulted in only 2.9% shared variance. Correlations were not statistically significant (p > .05) when controlling for the family-wise error rate with Bonferroni correction (Dunn, 1961).

## Psychopathology and Achievement Test Comparisons

As with the measures of intelligence and as hypothesized, all correlations between the achievement tests and the ASCA were low to near zero, ranging from -.11 to .10  $(Mdn_r = -.02)^3$ . The highest relationship between achievement and the ASCA indicated only 1.2% shared variance, and the correlations were not statistically significant (p > .05) when controlling for the family-wise error rate with Bonferroni correction (Dunn, 1961).

#### DISCUSSION

The present study examined the construct validity of the K-BIT, WISC-III, and ASCA through convergent and discriminant comparisons with a sample of students referred for special education evaluations. As hypothesized, the correlation between the K-BIT IQ Composite and WISC-III FSIQ was statistically significant and high with 79% shared variance, indicating the measurement of the same construct (general intelligence). The correlations between the K-BIT and WISC-III in the present study were nearly identical to and as high or higher than those obtained by Canivez (1995, 1996a, 1996b) and Grados and Russo-Garcia (1999). These correlations were also among the highest obtained on the K-BIT. Further, although the mean K-BIT IQ Composite and Vocabulary subtest standard scores were significantly higher than the mean WISC-III FSIQ and VIQ, respectively, the effect sizes were small and the roughly 2 IQ/standard score point differences were well within the standard errors of measurement for both measures and thus not of clinical importance. No other statistically significant mean differences were observed. These results are similar to those of Canivez (1995) where statistically significant but small differences were found, with the K-BIT Vocabulary subtest higher than the WISC-III VIQ

<sup>&</sup>lt;sup>2</sup>A table presenting all correlations between the Adjustment Scales for Children and Adolescents, the Wechsler Intelligence Scale for Children-Third Edition, and the Kaufman Brief Intelligence Test is available upon request.

<sup>&</sup>lt;sup>3</sup>A table presenting all correlations between the Adjustment Scales for Children and Adolescents and the achievement tests is available upon request.

but the K-BIT Matrices subtest lower than the WISC-III PIQ.

As in the Naugle et al. (1993) and Canivez (1995) studies where the K-BIT Vocabulary-Matrices discrepancy was compared to the WISC-III VIQ-PIQ discrepancy, the correlation in the present study was statistically significant but only low to moderate in magnitude (r = .38). As such, the K-BIT Vocabulary-Matrices discrepancy appears to be an inadequate predictor of the WISC-III VIQ-PIQ discrepancy, particularly if applied to individuals. As suggested by Canivez (1995), verbal-nonverbal differences may not have been in agreement because the Matrices subtest is a measure of fluid (Gf) abilities whereas the PIQ (and POI) may reflect Horn's Visual General Ability factor (Gv) or Carrol's Broad Visual Perception (GV) rather than fluid (Gf) abilities (Carroll, 1993a, 1993b; Sattler, 1992, 2001; Woodcock, 1990). It may also be a result of the Macmann and Barnett (1994) conclusion that the WISC-III measures only general intelligence (g) rather than a Verbal-Performance model and that "both the verbal and performance factors might be described (more logically and parsimoniously) as truncated or degraded versions of the general factor" (Macmann & Barnett, 1994, p. 180). Thus, nonverbal indices like the PIO and POI may simply be less reliable measures of general intelligence than VIQ and VCI. Factor-structure matrices presented by Macmann and Barnett demonstrated that WISC-III verbal subtests (Vocabulary, Information, Similarities, and Comprehension) loaded as well on the Performance factor as some performance subtests (Picture Completion and Picture Arrangement). In the present study, the K-BIT Vocabulary subtest was a much better predictor of the WISC-III VIQ (and VCI) than was the K-BIT Matrices subtest in predicting the WISC-III PIQ (and POI). In fact, the K-BIT Vocabulary subtest was correlated with the WISC-III PIQ and POI at higher levels than the K-BIT Matrices subtest (see Table 2). It can be concluded that the K-BIT and WISC-III both appear to be measuring the same general intelligence construct, as evidenced by the high correlation between the K-BIT IQ Composite and WISC-III FSIQ. With the recent publication of the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003) and its inclusion of a matrix reasoning subtest, future comparisons with the WISC-IV may produce higher correlations between the K-BIT Matrices subtest and WISC-IV POI as well as better agreement between verbal (crystallized) and nonverbal (fluid) discrepancies.

Comparisons of the K-BIT and WISC-III to the achievement measures produced correlations typical of comprehensive individually administered intelligence and achievement measures. With correlations ranging from .40 to .72 for the WISC-III ( $Mdn_r = .56$ ) and from .38 to .72 for the K-BIT ( $Mdn_r = .58$ ), these results are similar to those obtained by Canivez (1996a, 1996b) and Schultz (1997) and near the median intelligence-achievement correlation (r = .65) reported by Heath and Kush (1991) and the median intelligence-achievement correlation for small sample studies (r = .61) observed by Naglieri and Bornstein (2003). These correlations support the notion that although intelligence and achievement tests are positively related, they are indeed measuring different yet related constructs.

As McDermott (1995) found in comparing the ASCA and the DAS, the present study also found low to near zero correlations between the ASCA and the WISC-III and K-BIT. This was an expected finding because measures of intelligence (K-BIT and WISC-III) and measures of child psychopathology (ASCA) are theoretically very different constructs. Discriminant validity for the ASCA was further supported by the low to near zero correlations between the ASCA and the achievement measures.

Although this study is not a complete multitrait-multimethod matrix (Campbell & Fiske, 1959), it did simultaneously examine both convergent and discriminant validity within the same sample—something not yet presented in the empirical literature of the K-BIT, WISC-III, or ASCA. The correlations obtained in the present study are in the theoretically consistent direction and range for the types of comparisons made. Thus, there is evidence of construct validity for the K-BIT, WISC-III, and ASCA.

Although these results are positive, limitations in the present study affect generalizability of the results. First, most cases (n = 190) were obtained from rural school districts in Illinois (17 cases were from an urban area in Arizona), so the sample was not geographically representative of the United States. Further, generalizability is limited by the racial/ethnic diversity of the sample. Although African American and Hispanic/Latino students were included, their proportions were smaller than those of the population at large (although consistent with rural Illinois) and there were no Native American/American Indian or Asian American students. Another limitation is the fact that four different academic achievement tests were used, so specific ability-achievement correlations for each achievement test were not separately reported because this was beyond the scope and focus of the present study. A final limitation is that the majority of students in this study were learning disabled, and broad generalization to other groups is not recommended. Because the other disability groups had small sample sizes it was not possible to analyze data for these groups separately.

Overall, these results support the construct validity of the K-BIT, WISC-III, and ASCA. Additional research should also investigate comparisons of the K-BIT and WISC-III in measuring ability-achievement discrepancies in learning disability assessment (Canivez, 1996a, 1996b) as well as examine various VIQ–PIQ and Vocabulary–Matrices comparisons (Canivez, 1995). Replication of this study is needed and will help to further determine the validity of these clinically useful measures. Also, both the WISC-III and the K-BIT have recently been revised since these data were collected and results cannot be generalized to these newer measures. Data on these newer tests are presently being collected in a similar manner to investigate their construct validity.

- Achenbach, T. M. (1991). Manual for the Teacher's Report Form and 1991 Profile.
  Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T. M., & Edelbrock, C. (1983). Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: University of Vermont Department of Psychiatry.
- Allen, S. R., & Thorndike, R. M. (1995). Stability of the WAIS-R and WISC-III factor structure using cross-validation of covariance structures. *Journal of Clinical Psychology*, 51, 648-657.
- Algozzine, B., Eaves, R. C., Mann, L., & Vance, H. R. (1993). *The Slosson Full-Range Intelligence Test.* New York: Slosson Educational Publications.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- Axelrod, B. N., & Naugle, R. I. (1998). Evaluation of two brief and reliable estimates of the WAIS-R. International Journal of Neuroscience, 94, 85-91.
- Bell, N. L., Rucker, M., Finch, A. J., Jr., & Alexander, J. (2002). Concurrent validity of the Slosson Full-Range Intelligence Test: Comparisons with the Wechsler Intelligence Scale for Children-Third Edition and the Woodcock Johnson Tests of Achievement-Revised. *Psychology in the Schools, 39*, 31-38.
- Bowers, T. L., & Pantle, M. L. (1998). Shipley Institute for Living Scale and the Kaufman Brief Intelligence Test as screening instruments for intelligence. *Assessment*, 5, 187-195.
- Brown, L., Sherbenou, R. J., & Johnsen, S.
  K. (1990). Test of Nonverbal Intelligence-Revised. Austin, TX: Pro-Ed.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

- Canivez, G. L. (1995). Validity of the Kaufman Brief Intelligence Test: Comparisons with the Wechsler Intelligence Scale for Children-Third Edition. Assessment, 2, 101-111.
- Canivez, G. L. (1996a). Concurrent validity and diagnostic efficiency of the Kaufman Brief Intelligence Test in assessing severe discrepancy in reevaluations of students with learning disabilities. (ERIC Document Reproduction Service No. ED385959).
- Canivez, G. L. (1996b). Validity and diagnostic efficiency of the Kaufman Brief Intelligence Test in reevaluating students with learning disability. *Journal of Psychoeducational Assessment, 14, 4-19.*
- Canivez, G. L. (2001). Review of the Adjustment Scales for Children and Adolescents. In J. Impara and B. Plake (Eds.), The fourteenth mental measurements yearbook (pp. 22-24). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska.
- Canivez, G. L. (2004). Replication of the Adjustment Scales for Children and Adolescents core syndrome factor structure. *Psychology in the Schools, 41,* 191-199.
- Canivez, G. L., & Bordenkircher, S. E. (2002). Convergent and divergent validity of the Adjustment Scales for Children and Adolescents and the Preschool and Kindergarten Behavior Scales. Journal of Psychoeducational Assessment, 20, 30-45.
- Canivez, G. L., Perry, A. R., & Weller, E. M. (2001). Stability of the Adjustment Scales for Children and Adolescents. *Psychology in the Schools, 38*, 217-227.
- Canivez, G. L., & Rains, J. D. (2002). Construct validity of the Adjustment Scales for Children and Adolescents and the Preschool and Kindergarten Behavior Scales: Convergent and Divergent Evidence. *Psychology in the Schools*, *39*, 621-633.
- Canivez, G. L., & Sprouls, K. (in press). Assessing the construct validity of the

Adjustment Scales for Children and Adolescents. Journal of Psychoeducational Assessment.

- Canivez, G. L., & Watkins, M. W. (1998). Long term stability of the Wechsler Intelligence Scale for Children-Third Edition. *Psychological Assessment*, 10, 285-291.
- Canivez, G. L., & Watkins, M. W. (1999). Long term stability of the Wechsler Intelligence Scale for Children-Third Edition among demographic subgroups: Gender, race, and age. Journal of Psychoeducational Assessment, 17, 300-313.
- Canivez, G. L., & Watkins, M. W. (2001). Long term stability of the Wechsler Intelligence Scale for Children-Third Edition among students with disabilities. School Psychology Review, 30, 361-376.
- Canivez, G. L., & Watkins, M. W. (2002). Interrater agreement for Syndromic Profile Classifications on the Adjustment Scales for Children and Adolescents. Assessment for Effective Intervention.
- Canivez, G. L., Watkins, M. W., & Schaefer, B. A. (2002). Interrater agreement of Discriminant Classifications for the Adjustment Scales for Children and Adolescents. *Psychology in the Schools, 39*, 375-384.
- Caravajal, H., Hayes, J. E., Lackey, K. L., & Rathke, M. L. (1993). Correlations between scores on the Wechsler Intelligence Scale for Children-III and the General Purpose Abbreviated Battery of the Stanford-Binet IV. *Psychological Reports*, 72, 1167-1170.
- Carroll, J. B. (1993a). Human cognitive abilities: A survey of factor-analytic studies. New York: Cambridge University Press.
- Carroll, J. B. (1993b). What abilities are measured by the WISC-III? Monograph Series of the Journal of Psychoeducational Assessment, Wechsler Intelligence Scale for Children: Third Edition, 134-143.
- Chin, C. E., Ledesma, H. M. L., Cirino, P. T., Sevcik, R. A., Morris, R. D., Frijters, J. C., & Lovett, M. W. (2001). Relation

between Kaufman Brief Intelligence Test and WISC-III scores of children with RD. *Journal of Learning Disabilities*, 34, 2-8.

- Cicchetti, D., & Toth, S. L. (1991). A developmental perspective on internalizing and externalizing disorders. In D. Cicchetti & S. L. Toth (Eds.), Internalizing and externalizing expressions of dysfunction: Rochester symposium on developmental psychopathology (Vol. 2, pp. 1-19). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.
- DiCerbo, K. E., & Barona, A. (2000). A convergent validity study of the Differential Ability Scales and the Wechsler Intelligence Scale for Children-Third Edition with Hispanic children. Journal of Psychoeducational Assessment, 18, 344-352.
- Donders, J. (1995). Validity of the Kaufman Brief Intelligence Test (K-BIT) in children with traumatic brain injury. Assessment, 2, 219-224.
- Dumont, R., Cruse, C. L., Price, L., & Whelley, P. (1996). The relationship between the Differential Ability Scales (DAS) and the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) for students with learning disabilities. *Psychology in the Schools, 33*, 203-209.
- Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56, 52-64.
- Eisenstein, N., & Engelhart, C. J. (1997). Comparison of the K-BIT with short forms of the WAIS-R in a neuropsychological population. *Psychological Assessment*, 9, 57-62.
- Elliott, C. D. (1990). Differential Ability Scales; Introductory and technical handbook. San Antonio: The Psychological Corporation.
- Glass, G. V., & Hopkins, K. D. (1996). Statistical methods in education and psychology (3rd ed.). Needham Heights, MA: Allyn and Bacon.
- Grados, J. J., & Russo-Garcia, K. A. (1999). Comparison of the Kaufman Brief

Intelligence Test and the Wechsler Intelligence Scale for Children-Third Edition in economically disadvantaged African American youth. *Journal of Clinical Psychology*, 55, 1063-1071.

- Gresham, F. M., & Witt, J. C. (1997). Utility of intelligence tests for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. School Psychology Quarterly, 12, 249-267.
- Grice, J. W., Krohn, E. J., & Logerquist, S. (1999). Cross-validation of the WISC-III factor structure in two samples of children with learning disabilities. *Journal* of Psychoeducational Assessment, 17, 236-248.
- Hayes, S. C. (1999). Comparison of the Kaufman Brief Intelligence Test and the Matrix Analogies Test-Short Form in an adolescent forensic population. *Psychological Assessment*, 11, 108-110.
- Hays, J. R., Reas, D. L., & Shaw, J. B. (2002). Concurrent validity of the Wechsler Abbreviated Scale of Intelligence and the Kaufman Brief Intelligence Test among psychiatric inpatients. *Psychological Reports*, 90, 355-359.
- Heath, C. P., & Kush, J. C. (1991). Use of discrepancy formulas in the assessment of learning disabilities. In J. E. Obrzut & G. W. Hynd (Eds.), Neuropsychological foundations of learning disabilities: A handbook of issues, methods, and practice (pp. 287-307). San Diego, CA: Academic Press.
- Jensen, J. A., & Armstrong, R. J. (1985). Slosson Intelligence Test for children and adults: Expanded norms tables application and development. East Aurora: Slosson Educational Publications.
- Kamphaus, R. W., Petoskey, M. D., & Rowe, E. W. (2000). Current trends in psychological testing of children. *Professional Psychology: Research and Practice*, 31, 155-164.
- Kaufman, A. S. (1994). Intelligent testing with the WISC-III. New York: Wiley.
- Kaufman, A. S., & Kaufman, N. L. (1983). Administration and scoring manual for the

Kaufman Assessment Battery for Children. Circle Pines, MN: American Guidance Service.

- Kaufman, A. S., & Kaufman, N. L. (1990). Manual for the Kaufman Brief Intelligence Test. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S., & Kaufman, N. L. (1993). Kaufman Adolescent and Adult Intelligence Test manual. Circle Pines, MN: American Guidance Service.
- Kaufman, J. C., & Kaufman, A. S. (2001). Time for the changing of the guard: A farewell to short forms of intelligence tests. *Journal of Psychological Assessment*, 19, 245-267.
- Keith, T. Z., & Witta, E. L. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC-III: What does it measure? School Psychology Quarterly, 12, 89-107.
- Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. Psychological Assessment, 5, 395-399.
- Konold, T. R., Kush, J. C., & Canivez, G. L. (1997). Factor replication of the WISC-III in three independent samples of children receiving special education. *Journal of Psychoeducational Assessment*, 15, 123-137.
- Kush, J. C. (1996). Factor structure of the WISC-III for students with learning disabilities. *Journal of Psychoeducational* Assessment, 14, 32-40.
- Kush, J. C., & Watkins, M. W. (1997). Construct validity of the WISC-III verbal and performance factors for Black special education students. Assessment, 4, 297-304.
- Kush, J. C., Watkins, M. W., Ward, T. J., Ward, S. B., Canivez, G. L., & Worrell, F. C. (2001). Construct validity of the WISC-III for White and Black students from the WISC-III standardization sam-

ple and for Black students referred for psychological evaluation. *School Psychology Review*, *30*, 70-88.

- Landau, S., Milich, R., & Widiger, T. A. (1991). Predictive power methods may be more helpful for making a diagnosis than sensitivity and specificity. *Journal of Child & Adolescent Psychopharmacology*, 1, 343-351.
- Law, J. G., Jr., & Faison, L. (1996). WISC-III and KAIT results on adolescent delinquent males. *Journal of Clinical Psychology*, 52, 699-703.
- Levinson, E. M., & Folino, L. (1994). The relationship between the WISC-III and the Kaufman Brief Intelligence Test with students referred for gifted education. *Special Services in Schools*, *8*, 155-159.
- Livingston, R. B., Gray, R. M., Broquie, J. A. P., Dickson, T. L., Collins, A. L., & Spence, S. (2001). Construct validity of Wechsler factor scores. *Perceptual & Motor Skills*, 93, 78-80.
- Logerquist-Hansen, S., & Barona, A. (1994, August). Factor structure of the Wechsler Intelligence Scale for Children-III for Hispanic and Non-Hispanic White children with learning disabilities. (ERIC Document Reproduction Service No. ED379880).
- Lukens, J., & Hurrell, R. M. (1996). A comparison of the Stanford-Binet IV and the WISC-III with mildly retarded children. *Psychology in the Schools, 33,* 24-27.
- Macmann, G. M., & Barnett, D. W. (1994). Structural analysis of correlated factors: Lessons from the verbal-performance dichotomy of the Wechsler scales. *School Psychology Quarterly*, 9, 161-197.
- McDermott, P. A. (1988). Agreement among diagnosticians or observers: Its importance and determination. *Professional School Psychology*, 3, 225-240.
- McDermott, P. A. (1993). National standardization of uniform multisituational measures of child and adolescent behavior pathology. *Psychological Assessment*, 5, 413-424.

- McDermott, P. A. (1994). National profiles in youth psychopathology: Manual of Adjustment Scales for Children and Adolescents. Philadelphia: Edumetric and Clinical Science.
- McDermott, P. A. (1995). Sex, race, class, and other demographics as explanations for children's ability and adjustment: A national appraisal. *Journal of School Psychology, 33*, 75-91.
- McDermott, P. A., Marston, N. C., & Stott, D. H. (1993). Adjustment Scales for Children and Adolescents. Philadelphia: Edumetric and Clinical Science.
- McDermott, P. A., Watkins, M. W., Sichel, A. F., Weber, E. M., Keenan, J. T., Holland, A. M., & Leigh, N. M. (1995). The accuracy of new national scales for detecting emotional disturbance in children and adolescents. *The Journal of Special Education, 29*, 337-354.
- Merrell, K. W. (1994a). Preschool and Kindergarten Behavior Scales. Brandon, VT: CPPC.
- Merrell, K. W. (1994b). Assessment of behavioral social, and emotional problems. New York: Longman.
- Merrell, K. W. (2003). Behavioral, social, and emotional assessment of children and adolescents (2nd. ed.). Mahwah, NJ: Erlbaum.
- Naglieri, J. A. (1985a). Matrix Analogies Test-Short Form. San Antonio, TX: The Psychological Corporation.
- Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psychoeducational Assessment, 21, 244-*260.
- Naugle, R. I., Chelune, G. J., & Tucker, G. D. (1993). Validity of the Kaufman Brief Intelligence Test. *Psychological* Assessment, 5, 182-186.
- Prewett, P. N. (1992a). The relationship between the Kaufman Brief Intelligence Test (K-BIT) and the WISC-R with referred students. *Psychology in the Schools, 29*, 25-27.
- Prewett, P. N. (1992b). The relationship between the Kaufman Brief Intelligence Test (K-BIT) and the WISC-R

with incarcerated juvenile delinquents. Educational and Psychological Measurement, 52, 997-982.

- Prewett, P. N. (1995). A comparison of two screening tests (the Matrix Analogies Test-Short Form and the Kaufman Brief Intelligence Test) with the WISC-III. *Psychological Assessment*, 7, 69-72.
- Prewett, P. N., & Matavich, M. (1994). A comparison of referred students' performance on the WISC-III and the Stanford-Binet Intelligence Scale: Fourth Edition. *Journal of Psycho*educational Assessment, 12, 42-48.
- Prewett, P. N., & McCaffery, L. K. (1993). A comparison of the Kaufman Brief Intelligence Test (K-BIT), the Stanford-Binet Intelligence Scale (4th ed.), a 2subtest short form, and the Kaufman Test of Educational Achievement (K-TEA). Psychology in the Schools, 30, 299-304.
- Public Law (P.L.) 105-17. Individuals with Disabilities Education Act Amendments of 1997. (20 U.S.C. 1400 et seq.).
- Quay, H. C. (1986). Classification. In H. C. Quay & J. S. Werry (Eds.), *Psychopathological disorders of childhood* (3rd ed., pp. 1-42). New York: Wiley
- Raven, J. C. (1956). Coloured Progressive Matrices. London: H. K. Lewis.
- Reynolds, C. R., & Ford, L. (1994). Comparative three-factor solutions of the WISC-III and WISC-R at 11 age levels between 6<sup>1</sup>/<sub>2</sub> and 16<sup>1</sup>/<sub>2</sub> years. Archives of Clinical Neuropsychology, 9, 553-570.
- Reynolds, C. R., & Kamphaus, R. W. (1992). Behavior Assessment System for Children. Circle Pines, MN: American Guidance Service.
- Reynolds, C. R., & Kamphaus, R. W. (2004). Behavior Assessment System for Children (2nd ed.). Circle Pines, MN: American Guidance Service.
- Roid, G. H., Prifitera, A., & Weiss, L. G. (1993). Replication of the WISC-III factor structure in an independent sample. In B. A. Bracken & R. S. McCallum (Eds.), Weschler Intelligence Scale for Children: Third Edition, Journal of

Psychoeducational Assessment, Advances in Psychoeducational Assessment (pp. 6-21). Brandon, VT: CPPC.

- Roid, G. H., & Worrall, W. (1997). Replication of the Wechsler Intelligence Scale for Children-Third Edition four-factor model in the Canadian normative sample. *Psychological Assessment, 9*, 512-515.
- Rust, J. O., & Lindstrom, A. (1996). Concurrent validity of the WISC-III and Stanford-Binet IV. *Psychological Reports*, 79, 618-620.
- Rust, J. O., & Yates, A. G. (1997). Concurrent validity of the Wechsler Intelligence Scale for Children-Third Edition and the Kaufman Assessment Battery for Children. *Psychological Reports, 80,* 89-90.
- Saklofske, D. H., Schwean, V. L., Yaculic, R. A., & Quinn, D. (1994). WISC-III and SB:FE performances of children with Attention Deficit Hyperactive Disorder. *Canadian Journal of School Psychology*, 10, 167-171.
- Salvia, J., & Ysseldyke, J. E. (1995). Assessment. Boston: Houghton Mifflin.
- Sattler, J. M. (1992). Assessment of children: WISC-III and WPPSI-R supplement. San Diego, CA: Author.
- Sattler, J. M. (2001). Assessment of children: Cognitive applications (4th ed.). San Diego, CA: Author.
- Scardapane, J. R. (1996). Factor structure of the WISC-III for unclassified, learning-disabled, and high-IQ groups. Unpublished doctoral dissertation, Hofstra University, New York.
- Schultz, M. K. (1997). WISC-III and WJ-R Tests of Achievement: Concurrent validity and learning disability identification. *Journal of Special Education*, 31, 377-386.
- Silverstein, A. B. (1990). Short forms of individual intelligence tests. Psycho-logical Assessment: A Journal of Consulting and Clinical Psychology, 2, 3-11.
- Slate, J. R., Graham, L. S., & Bower, J. (1996). Relationships of the WISC-R and K-BIT for an adolescent clinical sample. *Adolescence*, 31, 777-782.

- Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psycho*educational Assessment, 12, 331-350.
- The Psychological Corporation (1992). Wechsler Individual Achievement Test. San Antonio, TX: Author.
- The Psychological Corporation (1999). Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: Author.
- The Psychological Corporation (2001). Wechsler Individual Achievement Test, Second Edition. San Antonio, TX: Author.
- Thompson, A., Browne, J., Schmidt, F., & Boer, M. (1997). Validity of the Kaufman Brief Intelligence Test and a four-subtest WISC-III short form with adolescent offenders. *Assessment*, 4, 385-394.
- Thorndike, R. L. (1992, April). Intelligence tests: What we have and what we should have. Paper presented at the annual convention of the National Association of School Psychologists, Nashville, TN.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). The Stanford-Binet Intelligence Scale: Fourth Edition. Chicago: Riverside.
- Trites, R. L., Blouin, A. G. A., & Laprade, K. (1982). Analysis of the Conners Teacher Rating Scale based on a large normative sample. *Journal of Consulting* and Clinical Psychology, 50, 615-623.
- Vo, D. H., Weisenberger, J. L., Becker, R., & Jacob-Timm, S. (1999). Concurrent validity of the KAIT for students in grade six and eight. *Journal of Psychoeducational Assessment*, 17, 152-162.
- Watkins, M. W., & Canivez, G. L. (1997). Interrater agreement of the Adjustment Scales for Children and Adolescents. *Diagnostique*, 22, 205-213.
- Watkins, M. W., & Canivez, G. L. (2001). Longitudinal factor structure of the WISC-III among students with disabilities. Psychology in the Schools, 38, 291-298.
- Watkins, M. W., Greenawalt, C. G., &

Marcell, C. M. (2002). Factor structure of the Wechsler Intelligence Scale for Children-Third Edition among gifted students. *Educational and Psychological Measurement*, 62, 164-172.

- Webber, L. S., & McGillivray, J. A. (1998). An Australian validation of the Kaufman Brief Intelligence Test (K-BIT) with adolescents with intellectual disability. Australian Psychologist, 33, 234-237.
- Wechsler, D. (1974). WISC-R manual. Wechsler Intelligence Scale for Children-Revised. New York: The Psychological Corporation.
- Wechsler, D. (1981). WAIS-R manual. Wechsler Adult Intelligence Scale-Revised. New York: The Psychological Corporation.
- Wechsler, D. (1991). Manual for the Wechsler Intelligence Scale for Children-Third Edition. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2003). Manual for the Wechsler Intelligence Scale for Children-Fourth Edition. San Antonio, TX: The Psychological Corporation.
- Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psycoeducational Assessment, 8,* 231-258.
- Woodcock, R. W., & Mather, N. (1989).
  WJ-R Tests of Achievement: Examiner's manual. In R. W. Woodcock & M. B. Johnson, Woodcock-Johnson Psycho-Educational Battery-Revised. Allen, TX: DLM Teaching Resources.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III Tests of Achievement. Itasca, IL: Riverside.
- Zachary, R. A. (1986). Shipley Institute of Living Scale: Revised manual. Los Angeles: Western Psychological Services.