ASSESSMENT
1995, Volume 2, Number 2, 101-111

VALIDITY OF THE KAUFMAN BRIEF INTELLIGENCE TEST: COMPARISONS WITH THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN—THIRD EDITION

Gary L. Canivez
Tempe School District #3
Tempe, Arizona

Concurrent validity of the Kaufman Brief Intelligence Test (K-BIT) with a sample of elementary- and middle-school students referred for multidisciplinary evaluations in a public school setting is presented. All correlations between the K-BIT and the Wechsler Intelligence Scale for Children–Third Edition (WISC-III) were significant. Correlations ranged from .36 (r^2 = .10) to .87 (r^2 = .75), M_r = .71 (M_r^2 = .50). K-BIT Vocabulary-Matrices discrepancy scores accounted for a significant but small proportion (13%) of the variability in WISC-III VIQ–PIQ discrepancies, but kappa (κ) coefficients for these discrepancies indicated that agreement was generally no better than chance. The K-BIT appears to be a promising general intellectual screening instrument when more comprehensive assessment is not possible or needed, but interpretation is best left at the IQ Composite level for the present time.

The Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990) is a new, individually administered test for providing a quick measure of intellectual abilities, which takes approximately 15 to 30 minutes to administer. The K-BIT was developed for use in screening high-risk children who may require more comprehensive evaluations, screening for educational diagnoses, assessing job applicants for hiring or placement, and estimating intellectual abilities as part of an emotional or personality assessment where obtaining an intellectual profile is not a primary concern. It was also constructed for use in (a) estimating intellectual abilities of large numbers of individuals where longer tests would be impractical, (b) reassessing the intellectual status of individuals previously administered a comprehensive intelligence test, (c) providing a valid estimate of intellectual abilities where constraints of time prohibit more lengthy measures, and (d) for research purposes. Most importantly, the K-BIT was developed in response to the lack of adequately normed and psychometrically sound brief intellectual measures that assess more than one trait or ability (Kaufman & Kaufman, 1990).

Gary L. Canivez is a school psychologist, Tempe School District #3 in Tempe, AZ and adjunct associate professor, Northern Arizona University.

A portion of this study was presented at the 1994 annual convention of the National Association of School Psychologists, Seattle, WA.

I thank W. Ken Newman and John Wilson for assistance in data collection and comments on the previous versions of this article. I also thank Joseph C. Kush, Marley W. Watkins, and Jennifer Hankins for comments on previous versions of this article.

Correspondence concerning this article should be addressed to Gary L. Canivez, Tempe School District #3, 3205 South Rural Road, P.O. Box 27708, Tempe, AZ 85285. For a copy of the *Excel™* for the Macintosh template created to calculate diagnostic efficiency statistics, send a preformatted Macintosh 3.5 inch disk with a self-addressed, stamped envelope and $5.00 to cover expenses.

Concurrent validity has been supported by comparisons between the K-BIT and the Test of Nonverbal Intelligence (TONI; Brown, Sherbeno, & Johnsen, 1990) and Slosson Intelligence Test (SIT; Jensen & Armstrong, 1985) (Kaufman & Kaufman, 1990). Correlation coefficients indicated good support for concurrent validity with the 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) for normal samples. K-BIT IQ Composite scores correlated between .58 and .69 (M_r = .63) with the K-ABC Mental Processing Composite across three age ranges. Correlations between the K-BIT IQ Composite and WISC-R Full Scale IQ (r = .80) and WAIS-R Full Scale IQ (r = .75) also supported the validity of the K-BIT. In addition, the K-BIT Vocabulary subtest correlated better with the WISC-R and WAIS-R Verbal IQ than Performance IQ (as would be expected) while the K-BIT Matrices subtest correlated *equally* well with the WISC-R and WAIS-R Verbal IQ and Performance IQ (Kaufman & Kaufman, 1990). Prewett (1992a, 1992b) has also found significant correlations between the K-BIT and WISC-R for samples of referred students and incarcerated juvenile delinquents, whereas Naugle, Chelune, and Tucker (1993) found significant correlations with the WAIS-R for a sample of patients receiving neuropsychological evaluations.

The K-BIT Matrices subtest was included to assess fluid/nonverbal ability, whereas the Vocabulary subtests (Expressive Vocabulary and Definitions) assess crystallized/verbal ability. Horn and Cattell (1966) refer to crystallized abilities as skills heavily dependent upon cultural experiences or direct educational instruction, whereas fluid abilities are reflected in culture-fair tasks and allow one to adapt and function when faced with novel or unfamiliar problems. By including *both* verbal and nonverbal subtests (most brief intelligence tests focus on one or the other), the K-BIT is able to measure two very different skill areas which allow the examiner to assess verbal-nonverbal discrepancies as is done with the Wechsler scales (Kaufman & Kaufman, 1990). Kaufman and Kaufman are cautious about the interpretation of significant (α = .05 or .01) Vocabulary–Matrices discrepancies and recommend 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, which should be further investigated with a comprehensive assessment.

Unfortunately, to date, there appears to be only one study, Naugle et al. (1993), investigating the relationship between discrepancies obtained between the Vocabulary and Matrices subtests of the K-BIT and similar discrepancies on a comprehensive intellectual measure. The tenability of hypotheses regarding the Vocabulary–Matrices discrepancy predicting similar discrepancies in comprehensive measures of intelligence is an empirical question requiring study. Studies referenced in the K-BIT manual (Kaufman & Kaufman, 1990) comparing the K-BIT to the WISC-R and the WAIS-R did not appear to examine the relationship between the K-BIT Vocabulary–Matrices discrepancy and WISC-R or WAIS-R Verbal IQ–Performance IQ (VIQ–PIQ) discrepancy. Prewett's (1992a, 1992b) analyses comparing the K-BIT to the WISC-R also did not examine such discrepancies. Naugle et al. found correlations between the Vocabulary–Matrices discrepancy and the WAIS-R VIQ–PIQ discrepancy ranged from .23 to .59 across eight age ranges, while the correlation for the total sample was .46. If the K-BIT Vocabulary–Matrices discrepancy is to be a useful comparison, it must show acceptable concurrent and predictive validity in relation to other measures assessing verbal and nonverbal discrepancies.

One year following the publication of the K-BIT, the WISC-III (Wechsler, 1991) was published as a revision of the WISC-R, and reflects updating of materials as well as a new standardization sample closely approximating the characteristics of 1988

United States Census data on selected demographic variables. Although the K-BIT manual presents studies examining the relationship between the K-BIT and WISC-R, to date, no published studies were available that examined the relationship between the K-BIT and the WISC-III for normal or clinical populations. National surveys examining assessment practices of school psychologists indicated that the Wechsler scales are the most frequently used measures of general intelligence (Goh, Teslow, & Fuller, 1981; Stinnett, Havey, & Oehler-Stinnett, 1994), thus, assessing the relationship between the K-BIT and WISC-III is important to practitioners. The present study examined the relationship between the K-BIT Vocabulary, Matrices, and IQ Composite standard scores and the WISC-III Verbal, Performance, and Full Scale IQs with students referred for psychoeducational evaluations in a public-school setting. This study also examined relations between the K-BIT and WISC-III Verbal Comprehension (VCI), Perceptual Organization (POI), Freedom from Distractibility (FDI), and Processing Speed (PSI) Index Scores. It was predicted that the K-BIT IQ Composite would correlate significantly with the WISC-III FSIQ and that the K-BIT Vocabulary subtest would have higher correlations with the WISC-III VIQ and VCI than PIQ and POI, whereas the K-BIT Matrices subtest would have higher correlations with the WISC-III PIQ and POI than VIQ and VCI. Lowest correlations were hypothesized to be with FDI and PSI scores that are composed of subtests, which, based upon factor analytic data, are not as highly related to VC or PO (Wechsler, 1991). The present study also examined the concurrent validity and level of agreement between the K-BIT Vocabulary-Matrices discrepancies and WISC-III VIQ-PIQ discrepancies at various significance levels.

Method

Subjects

The 137 subjects in the present study were elementary- (K–6th grade) and middle-school (6th–8th grade) students in a major, southwest metropolitan public-school system, who were referred for initial or triennial multidisciplinary evaluation. Sixty-six percent (n = 91) were male, 34% (n = 46) were female, and the mean age of the subjects was 11.7 years (SD = 2.15, range = 6–15 years). Ethnic characteristics of the subjects were as follows: Caucasian, 45% (n = 62); Black, 11% (n = 15); Hispanic, 34% (n = 46); Native American, 9% (n = 12); and Hispanic-Native American, 1% (n = 2). All subjects in this study were sufficiently proficient in English to enable appropriate administration of present tests, although some were bilingual. Bilingual subjects were evaluated by a bilingual school psychologist. Seventy-six percent (n = 104) were monolingual English speakers, 15% (n = 20) had primary language of English and secondary language of Spanish, and 9% (n = 13) had primary language of Spanish and secondary language of English. Thirty-three (24%) of the evaluations were initial evaluations, whereas 104 (76%) were triennial reevaluations. Results from these evaluations indicated that 22 (16%) subjects were not disabled, whereas 95 (69%) were learning disabled, 5 (4%) were seriously emotionally disabled, 13 (9%) were mildly mentally retarded, 1 (1%) was moderately mentally retarded, and 1 (1%) was speech/language impaired. State special-education rules and regulations used for classification of students into the above categories were similar to those specified by the United States Department of Education (1992). Learning disability was operationalized as a severe discrepancy between ability and achievement using a regression approach (Reynolds, 1984), and 1.5 standard errors of estimate was suggested as a criterion for severe discrepancy. Mental-retardation classification required significant deficits in both intellectual abilities and adaptive behavior, whereas emotional disabilities was defined as one of five emotional characteristics that adversely impacted the student's educational performance. Two of the students who were not disabled were identified through evaluation as intellectually gifted.

Measures

KBIT

"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 composed 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. 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 (M_r = .94), .89 to .98 (M_r = .93), and .74 to .95 (M_r = .88), respectively. Test-retest stability estimates for the IQ Composite, Vocabulary, and Matrices scores with four age samples ranged from .92 to .95 (M_r = .94), .86 to .97 (M_r = .94), and .80 to .92 (M_r = .85), respectively (Kaufman & Kaufman, 1990).

WISC-III

The WISC-III is an individually administered test of intellectual abilities for children aged 6 years through 16 years, 11 months (Wechsler, 1991). As with previous editions, the WISC-III is composed of several subtests that measure different aspects of intelligence and yields three composite IQs (viz., VIQ, PIQ, and FSIQ), which provide estimates of the individual's verbal, nonverbal, and general intellectual abilities. The WISC-III also yields four optional factor-based index scores (viz., VCI, POI, FDI, and 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. Concurrent validity studies generally found moderately high correlations with other intellectual ability measures and VIQ tended to correlate higher with verbal ability measures than with nonverbal ability measures, whereas PIQ tended to correlate higher with nonverbal ability measures than with verbal ability measures (Wechsler, 1991) as expected.

Procedure

Subjects were administered the K-BIT and WISC-III in counterbalanced order, during the same test session, as part of a comprehensive psychoeducational evaluation to determine exceptionality. Evaluations were conducted by three licensed and nationally certified school psychologists. K-BIT Vocabulary, Matrices, and IQ Composite standard scores were obtained, and Vocabulary-Matrices discrepancy scores were examined for determination of significant differences at the α = .05 and α = .01 levels (see Table C.5, Kaufman & Kaufman, 1990). Vocabulary-Matrices discrepancy scores were also examined for determination of "abnormality" based upon a 5% population prevalence criterion (see Table 3.2, Kaufman & Kaufman, 1990).

WISC-III VIQ, PIQ, FSIQ, VCI, POI, FDI, and PSI scores were also obtained; and of the 137 subjects, 1 was not administered the Digit Span subtest and 6 were not administered the Symbol Search subtest. Thus, analyses for the FDI and PSI are based on ns = 136 and 131, respectively. VIQ-PIQ discrepancy scores were examined for determination of significant differences at the α = .05 level (see Table B.1, Wechsler, 1991, p. 261) and α = .01 level. Critical values for VIQ-PIQ significance at the α = .01 level are not available in the WISC-III manual; and, although Naglieri (1993) provided values for significant VIQ-PIQ differences (α = .01), these values are inflated due to Bonferroni correction, which adjusts for the family-wide error rate in multiple discrepancy comparisons. The present study examined only one WISC-III pairwise comparison (viz., VIQ-PIQ), so critical values for significance of α = .01 were obtained following the formula: Difference score = $z(SEM_a^2 + SEM_b^2)^{1/2}$, where z = 2.5758 (value from the normal curve corresponding to α = .01), SEM_a = standard error of measurement for VIQ at the appropriate age level, and SEM_b = standard error of measurement for PIQ at the appropriate age level (Anastasi, 1988; Guilford & Fruchter, 1978). The SEMs used for each age level were obtained from Table 5.2 in the WISC-III manual (Wechsler, 1991). VIQ–PIQ discrepancies were also examined with regard to the 5% population prevalence criterion level (see Table B.2, Wechsler, 1991).

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. In addition, the K-BIT Vocabulary–Matrices discrepancy score was used as a predictor (continuous independent variable) of the WISC-III VIQ–PIQ discrepancy score in a linear regression analysis. Diagnostic efficiency statistics were calculated as recommended by Kessel and Zimmerman (1993) and automated by Canivez and Watkins (in press) to further evaluate the K-BIT Vocabulary–Matrices discrepancy. Kappa (κ) coefficients (Cohen, 1960) were calculated to assess the level of agreement between Vocabulary–Matrices and VIQ–PIQ discrepancies at the $\alpha = .05$ and .01 levels, as well as with the 5% population prevalence criterion. To test whether κ coefficients were significant, z-tests were performed as recommended by Fleiss (1981).

Results

Pearson product-moment correlation coefficients and r^2s for the K-BIT and WISC-III are presented in Table 1. All correlations were significant ($p < .0001$). Correlations ranged from .87 to .36 with a mean correlation of .71.[1] Differences between correlation coefficients were tested using Hotelling's formula for a t-test (Guilford & Fruchter, 1978). As expected, the Vocabulary subtest had a significantly higher correlation with the WISC-III VIQ than with PIQ, $t(134) = 5.01$, $p < .0001$, and significantly higher correlation with VCI than with POI, $t(134) = 4.47$, $p < .0001$. The Matrices subtest correlated equally well with PIQ, POI, VIQ, and VCI, as no significant differences were noted among these correlations. These results are consistent with those obtained in other studies (Kaufman & Kaufman, 1990; Prewett, 1992a, 1992b; Naugle et al., 1993). The lowest correlations obtained were with the PSI, as expected.

To investigate the relationship between the K-BIT and WISC-III for different levels of intellectual abilities, the sample was divided into three subgroups (FSIQ < 70, 70 ≤ FSIQ ≤ 84, and 85 ≤ FSIQ ≤ 114) corresponding to scores below $-2SD$ ($n = 29$), between $-2SD$ and $-1SD$ ($n = 55$), and between $-1SD$ and $+1SD$ ($n = 49$), respectively. There was an insufficient number of subjects with FSIQs above 115 for comparison. Correlations between the K-BIT IQ Composite and WISC-III FSIQ for these subgroups were .59, .60, and .63, respectively. These correlations were not significantly different and indicated that the K-BIT was equally valid within each of these IQ ranges. These correlations are smaller than those obtained for the total sample due to restricted ranges, but were still moderately high.

Descriptive statistics are presented in Table 2. Subjects obtained equivalent K-BIT IQ Composite and WISC-III Full Scale IQ scores, $t(136) = 1.08$, *ns*. However, subjects scored significantly higher on the K-BIT Vocabulary subtest than on the WISC-III VIQ, $t(136) = 4.53$, $p < .0001$; whereas subjects scored significantly lower on the K-BIT Matrices subtest than on the WISC-III PIQ, $t(136) = 2.11$, $p < .05$. Although significant, these mean differences were not large or of practical significance, as they are well within the standard errors of measurement for both measures. Naugle et al. (1993) also reported significant, but small, differences with subjects scoring consistently *higher* on the K-BIT than on the WAIS-R.

The regression analysis assessing the ability of the K-BIT Vocabulary–Matrices discrepancy score to predict the WISC-III VIQ–PIQ discrepancy score was significant, $F(1, 135) = 19.27$, $p < .0001$. However, only 13% ($r^2 = .13$) of the variability in WISC-III VIQ–PIQ discrepancy ($M_{\text{VIQ-PIQ}} = -10.04$, $SD = 11.16$) was accounted for by the K-BIT Vocabulary–Matrices discrepancy ($M_{\text{V-M}} = -4.28$, $SD = 14.98$). Naugle et al. (1993) found the K-BIT Vocabulary–Matrices discrepancy accounted for only 21% of the variability in WAIS-R VIQ–PIQ discrepancy ($r^2 = .21$). VIQ–PIQ

[1] The average correlation coefficient was obtained using Fisher's Z transformation (Guilford & Fruchter, 1978).

Table 1
Pearson Product-Moment Correlation Coefficients for the K-BIT and WISC-III

	K-BIT		
WISC-III	Vocabulary	Matrices	IQ Composite
VIQ	.80 (.64)	.67 (.45)	.84 (.71)
PIQ	.62 (.38)	.74 (.55)	.79 (.62)
FSIQ	.76 (.58)	.75 (.56)	.87 (.76)
VCI	.79 (.62)	.66 (.44)	.83 (.69)
POI	.62 (.38)	.70 (.49)	.76 (.58)
FDI	.65 (.42)	.60 (.36)	.71 (.50)
PSI	.36 (.13)	.47 (.22)	.48 (.23)

Note. r^2s presented in parentheses. All correlations significant $p < .0001$. K-BIT = Kaufman Brief Intelligence Test; 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.

Table 2
Descriptive Statistics for K-BIT and WISC-III Scores

Variable	*n*	*M*	*SD*	Range
K-BIT				
Vocabulary	137	81.94	14.08	52 - 123
Matrices	137	86.23	16.18	40 - 130
IQ Composite	137	82.39	14.46	49 - 118
WISC-III				
VIQ	137	78.31	15.18	46 - 124
PIQ	137	88.35	16.55	48 - 141
FSIQ	137	81.66	15.89	48 - 130
VCI	137	79.20	14.96	50 - 123
POI	137	89.01	17.55	50 - 135
FDI	136	80.68	12.91	50 - 115
PSI	131	91.75	14.28	58 - 137

Note. K-BIT = Kaufman Brief Intelligence Test; 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.

discrepancies were also significantly larger than Vocabulary–Matrices discrepancies $t(136) = 4.44$, $p < .0001$. Table 3 presents the frequency data for students demonstrating various K-BIT Vocabulary–Matrices discrepancies and WISC-III VIQ–PIQ discrepancies for $\alpha = .05$ and .01 and for the 5% population prevalence levels. Table 4 presents the diagnostic efficiency statistics for these comparisons.

Only for $\alpha = .05$, was κ significant, indicating that agreement between K-BIT Vocabulary–Matrices discrepancy and WISC-III VIQ–PIQ was beyond chance. The magnitude of this agreement was, however, slight (Everitt & Hay, 1992) and not particularly meaningful in light of the high false positive and false negative rates. Three subjects, in fact, showed a significant K-BIT Vocabulary–Matrices discrepancy but demonstrated a significant WISC-III VIQ–PIQ discrepancy (opposite of predicted direction). For $\alpha = .01$ and the 5% population prevalence level, κ coefficients were not significant and represented chance levels of agreement between Vocabulary–Matrices and VIQ–PIQ discrepancies. Kappa coefficients indicated that there was little to no agreement between the K-BIT Vocabulary–Matrices and WISC-III VIQ–PIQ discrepancies.

Discussion

The present study examined the concurrent validity of the K-BIT with the WISC-III in a sample of elementary- and middle-school students referred for multidisciplinary evaluations. The K-BIT IQ Composite, Vocabulary, and Matrices scores compared favorably to the WISC-III IQ and Index scores. These data provide evidence supporting the concurrent validity of the K-BIT as a brief estimate of general intellectual abilities. As such, it is a substantial improvement over the poorly normed and psychometrically inferior Slosson Intelligence Test (Kaufman, 1990; Oakland, 1985; Reynolds, 1985). Because the K-BIT is composed of both verbal and nonverbal subtests, it is also an improvement over instruments such as the Matrix Analogies Test (short and expanded forms; Naglieri, 1985a, 1985b), Test of Nonverbal Intelligence–Revised (Brown, Sherbenou, & Johnsen, 1990), and Peabody Picture Vocabulary Test–Revised (Dunn & Dunn, 1981).

K-BIT Vocabulary, Matrices, and IQ Composite and WISC-III VIQ, PIQ, FSIQ, VCI, and POI correlations obtained in the present study were slightly lower than correlations between WISC-III Vocabulary and Block Design with VIQ, PIQ, FSIQ, VCI, and POI as presented in the WISC-III manual (Wechsler, 1991, p. 281). These two subtests (Vocabulary and Block Design) are frequently combined in a two subtest short form used for intellectual screening purposes (Kaufman, 1990; Sattler, 1992). However, 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. While retaining high correlations with the various WISC-III IQ and Index scores, the K-BIT has advantages of motor-free subtests, and a psychologist is not required for administration. All WISC-III short forms require a psychologist for administration. Another problem with short forms is that they are developed utilizing standardization data in which the subjects were administered the entire test, and the resulting scores may not correspond if only the short-form subtests were administered in isolation. Regardless of advantages, use of brief intellectual measures (K-BIT or short forms) has not been recommended when making educational or diagnostic decisions (Kaufman, 1990; Sattler, 1992; Silverstein, 1990). Kaufman and Sattler discuss the loss of information related to profile analysis when short forms or intellectual screening instruments are used; however, some would argue that there is a lack of empirical support for profile or ipsative analysis among comprehensive intellectual ability measures (Hale, 1979; Hale & Landino, 1981; Hale & Saxe, 1983; McDermott, Fantuzzo, & Glutting, 1990; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992; McDermott, Glutting, Jones, & Noonan, 1989; McDermott, Glutting, Jones, Watkins, & Kush, 1989; Watkins & Kush, 1994). The diagnostic utility of the K-BIT in clinical decision-making has yet to be investigated and remains an open empirical question.

Table 3
Frequencies of Subjects Showing Significant (α = .05 and .01) and "Abnormal" (5% population prevalence) K-BIT Vocabulary–Matrices and WISC-III Verbal IQ–Performance IQ Discrepancies

		WISC-III VIQ–PIQ	
α = .05	*ns*	VIQ > PIQ	PIQ > VIQ
K-BIT Vocabulary–Matrices			
ns	43_d	1_c	30_c
Vocabulary > Matrices	14_b	1_a	3
Matrices > Vocabulary	18_b	0	27_a
		WISC-III VIQ–PIQ	
α = .01	*ns*	VIQ > PIQ	PIQ > VIQ
K-BIT Vocabulary–Matrices			
ns	69_d	1_c	30_c
Vocabulary > Matrices	11_b	0_a	0
Matrices > Vocabulary	15_b	0	11_a
		WISC-III VIQ–PIQ	
5% Population Prevalence	5%	VIQ > PIQ	PIQ > VIQ
K-BIT Vocabulary–Matrices			
> 5%	112_d	0_c	13_c
Vocabulary > Matrices	3_b	0_a	0
Matrices > Vocabulary	7_b	0	2_a

Note. Numbers along the diagonal (moving from the upper left to the lower right and subscripted with a or d) indicate consistent results and agreement between K-BIT Vocabulary–Matrices discrepancy and WISC-III VIQ–PIQ discrepancy. False negatives fall above the diagonal while false positives fall below the diagonal. Subscripts a, b, c, and d correspond to the appropriate cells in a 2 × 2 diagnostic efficiency statistics table presented in Kessel and Zimmerman (1993). K-BIT = Kaufman Brief Intelligence Test; WISC-III = Wechsler Intelligence Scale for Children-Third Edition; VIQ = Verbal IQ; PIQ = Performance IQ; *ns* = not significant.

The present study indicates that the K-BIT falls short of its goal of assessing a verbal-nonverbal dichotomy, in that there was little to no agreement between K-BIT Vocabulary–Matrices discrepancies and WISC-III VIQ–PIQ discrepancies. Given the small proportion of variability of WISC-III VIQ–PIQ differences accounted for by K-BIT Vocabulary–Matrices differences (13%), low sensitivity estimates, low positive predictive power, and the high false positive and false negative predictions from the K-BIT Vocabulary–Matrices discrepancy, clinicians should not use the K-BIT Vocabulary–Matrices discrepancy to make predictions of possible verbal-nonverbal differences in comprehensive intelligence tests such as the WISC-III (or WAIS-R) (Naugle et al., 1993). These data suggest that the K-BIT does not possess adequate sensitivity or positive predictive power to correctly identify subjects who have VIQ–PIQ discrepancies. This may be partly related to the fact that the K-BIT is composed of only two subtests and does not sample the respective domains as well as a comprehensive intellectual measure. It may also be due to the unreliability of difference (discrepancy) scores (Silverstein, 1981; Thorndike & Hagen, 1977).

In addition to problems associated with the K-BIT discussed previously, another possible cause for

Table 4
Diagnostic Efficiency Statistics for Agreement Between K-BIT Vocabulary–Matrices and WISC-III Verbal IQ–Performance IQ Discrepancies

	α = .05	α = .01	5% PP
Sensitivity	.47	.26	.13
Specificity	.57	.73	.92
Positive Predictive Power	.47	.30	.17
Negative Predictive Power	.58	.69	.90
False Positive Rate	.43	.27	.08
False Negative Rate	.53	.74	.87
Overall Correct Classification	.52	.58	.83
κ	.14	.05	.07
SE_{κ}	.07	.07	.08
z	1.99	.66	.91
p	.05	*ns*	*ns*

Note. K-BIT = Kaufman Brief Intelligence Test; WISC-III = Wechsler Intelligence Scale for Children-Third Edition; PP = Population Prevalence.

the low agreement between K-BIT and WISC-III verbal-nonverbal discrepancies has been suggested by Macmann and Barnett (1994). They concluded that the WISC-III measures only general intelligence (g) rather than a Verbal-Performance model and noted that "both the verbal and performance factors might be described (more logically and parsimoniously) as truncated or degraded versions of the general factor" (p. 180). Thus, nonverbal indexes like the PIQ and POI may simply be less reliable measures of general intelligence than VIQ and VCI. Factor-structure matrices presented by Macmann and Barnett demonstrate 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). This could also help to explain why the K-BIT Vocabulary subtest correlated higher with VIQ and VCI than PIQ and POI, but the Matrices subtest correlated equally well with the WISC-III VIQ, VCI, PIQ, and POI. This conjecture is supported by the highly correlated VIQ and PIQ (r = .76) and VCI and POI (r = .74) scores found in the present study.

Alternatively, verbal-nonverbal differences may not have been in agreement, as the Matrices subtest was designed to be a measure of fluid (G_f) abilities whereas the PIQ may be thought of as reflecting Horn's Visual General Ability factor (G_v), rather than fluid (G_f) abilities (Sattler, 1992; Woodcock, 1990). Regardless, at the present time, the K-BIT should be considered as only an estimate of general intelligence (g) until additional research can clarify the nature of and relationships between the Vocabulary and Matrices subtests with other instruments and populations hypothesized to reflect multiple intellectual factors.

Future research should continue to examine the K-BIT's relationship with other comprehensive intellectual ability measures, with different samples of normal individuals, and with those who have specific disabilities to further define its psychometric characteristics. Specifically, researchers should continue to examine the K-BIT's ability to differentiate verbal/crystallized and nonverbal/fluid abilities. Differences between racial or ethnic groups and bilingual subjects should also be explored to examine possible differential validity. Research should also examine clinical decision-making in using the K-BIT versus a comprehensive intellectual measure. A possible use of the K-BIT could be to reduce the assessment time during reevaluations.

If assessment decisions are similar when using the K-BIT compared to a comprehensive intellectual measure, then this substitution may be a cost- and time-effective practice that could free time for the clinician to engage in alternative assessment practices or provide alternative services. If future research replicates the present findings, the K-BIT will likely become a popular and frequently used test for quickly estimating the general intelligence of individuals.

References

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1990). *Test of Nonverbal Intelligence–Revised.* Austin, TX: Pro-Ed.

Canivez, G. L., & Watkins, M. W. (in press). Automated calculation of diagnostic efficiency statistics. *Behavior Research Methods, Instruments, & Computers.*

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

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

Everitt, B. S., & Hay, D. F. (1992). *Talking about statistics: A psychologist's guide to data analysis.* New York: Halsted.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Goh, D. S., Teslow, C. J., & Fuller, G. B. (1981). The practice of psychological assessment among school psychologists. *Professional Psychology, 12,* 696-706.

Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education* (6th ed.). New York: McGraw-Hill.

Hale, R. L. (1979). The utility of the WISC-R subtest scores in discriminating among adequate and underachieving children. *Multivariate Behavioral Research, 14,* 245-253.

Hale, R. L., & Landino, S. A. (1981). Utility of the WISC-R subtest analysis in discriminating among groups of conduct problem, withdrawn, mixed, and non-problem boys. *Journal of Consulting and Clinical Psychology, 41,* 91-95.

Hale, R. L., & Saxe, J. E. (1983). Profile analysis of the Wechsler Intelligence Scale for Children–Revised. *Journal of Psychoeducational Assessment, 1,* 155-162.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57,* 253-270.

Jensen, J. H., & Armstrong, R. J. (1985). *Slosson Intelligence Test for children and adults: Expanded norms tables application and development.* East Aurora, NY: Slosson Educational Publication.

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

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.

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.

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., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment, 8,* 290-302.

McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., & Baggaley, A. R. (1992). Illusions of meaning in the ipsative assessment of children's ability. *Journal of Special Education, 25,* 504-526.

McDermott, P. A., Glutting, J. J., Jones, J. N., & Noonan, J. V. (1989). Typology and prevailing composition of core profiles in the WAIS-R standardization sample. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 118-125.

McDermott, P. A., Glutting, J. J., Jones, J. N., Watkins, M. W., & Kush, J. (1989). Core profile types in the WISC-R national sample: Structure, membership, and applications. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 292-299.

Naglieri, J. A. (1985a). *Matrix Analogies Test-short form.* San Antonio, TX: The Psychological Corporation.

Naglieri, J. A. (1985b). *Matrix Analogies Test-expanded form.* San Antonio, TX: The Psychological Corporation.

Naglieri, J. A. (1993). Pairwise and ipsative comparisons of WISC-III IQ and Index scores. *Psychological Assessment, 5,* 113-116.

Naugle, R. I., Chelune, G. J., & Tucker, G. D. (1993). Validity of the Kaufman Brief Intelligence Test. *Psychological Assessment, 5,* 182-186.

Oakland, T. (1985). Review of Slosson Intelligence Test. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 1401-1403). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska.

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.

Reynolds, C. R. (1984). Critical measurement issues in learning disabilities. *Journal of Special Education, 18,* 1-26.

Reynolds, W. M. (1985). Review of Slosson Intelligence Test. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 1403-1404). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska.

Sattler, J. M. (1992). *Assessment of children* (3rd ed.). San Diego, CA: Author.

Silverstein, A. B. (1981). Reliability and abnormality of test score differences. *Journal of Clinical Psychology, 37,* 392-394.

Silverstein, A. B. (1990). Short forms of individual intelligence tests. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2,* 3-11.

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment, 12,* 331-350.

Thorndike, R. L., & Hagen, E. P. (1977). *Measurement and evaluation in psychology and education* (4th ed.). New York: Wiley.

United States Department of Education. (1992). Assistance to states for the education of children with disabilities program and preschool grants for children with disabilities. *Federal Register, 57,* 44794-44852.

Watkins, M. W., & Kush, J. C. (1994). Wechsler subtest analysis: The right way, the wrong way, or no way? *School Psychology Review, 23,* 640-651.

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

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

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

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231-258.