Journal of Psychoeducational Assessment
1996, *14*, 4-19

VALIDITY AND DIAGNOSTIC EFFICIENCY OF THE KAUFMAN BRIEF INTELLIGENCE TEST IN REEVALUATING STUDENTS WITH LEARNING DISABILITY

Gary L. Canivez
Eastern Illinois University

The present study examined the concurrent validity and diagnostic efficiency of the Kaufman Brief Intelligence Test (K-BIT). Seventy-five students with learning disability were administered the K-BIT, Wechsler Intelligence Scale for Children-Third Edition (WISC-III), and Woodcock-Johnson-Revised Tests of Achievement (WJ-R ACH) as part of their triennial reevaluation. Correlations between the K-BIT and the WISC-III ranged from .18 (r^2 = .03) to .82 (r^2 = .67), M_r = .62 ($M_r{}^2$ = .38). High levels of agreement were obtained between the K-BIT and WISC-III in identifying severe achievement-ability discrepancies on the WJ-R ACH. The K-BIT appears to be a promising general intellectual screening instrument when more comprehensive assessment is not possible or needed.

School psychologists report spending significant portions of time (half to two-thirds) in evaluation of students for possible placement in special education programs (Goh, Teslow, & Fuller, 1981; Hutton, Dubes, & Muir, 1992; Reschley, Genshaft, & Binder, 1987; Smith, 1984). With specific learning disability (SLD) becoming the category of special education with the highest proportion of students (Heath & Kush, 1991), much of a school psychologist's time is spent evaluating such students (Reschley et al., 1987). In the evaluation and identification of students with SLD it is necessary in part to identify the presence of "a severe discrepancy between achievement and intellectual ability" (United States Department of Education [USDE], 1992, p. 44823). Comprehensive intellectual ability measures are used most frequently to assess intellectual ability, and the Wechsler scales are by far the most frequently used by school psychologists (Goh et al., 1981; Hutton et al., 1992).

Gary L. Canivez is an assistant professor in the Department of Psychology at Eastern Illinois University.

The 75 students with learning disability in this study are a subset of students from a larger study (N = 137; Canivez, 1995) that addressed the concurrent validity of the K-BIT and that was presented at the 1994 annual convention of the National Association of School Psychologists, Seattle. A summary of these results was presented at the 1995 annual convention of the National Association of School Psychologists, Chicago.

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 and Marley W. Watkins for comments on previous versions of this article.

Correspondence should be addressed to Gary L. Canivez, Ph.D., Department of Psychology, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920-3099, cfglc@eiu.edu, http://www.eiu.edu/ac/sci/psy/CanivezHP. For a copy of the *Microsoft*® *Excel*™ templates created to calculate diagnostic efficiency statistics, send an *Apple*® *Macintosh*™ preformatted 3.5 inch disk with a self-addressed stamped envelope and $5.00 to cover expenses.

In assessing the presence of severe discrepancy between achievement and intellectual ability, experts in psychological measurement indicate that the technically appropriate method for determining severe discrepancy between achievement and intellectual ability is the use of a regression-based mathematical formula (Heath & Kush, 1991; Reynolds, 1984; Wilson & Cone, 1984) when the tests used were not co-normed on the same sample. This approach accounts for regression to the mean effects as well as measurement error. This approach also may assist in building consistency in identifying students with severe discrepancy and classification of learning disability (Ross, 1992).

Once identified as disabled, reevaluation is required "every three years, or more frequently if conditions warrant" (USDE, 1992, p. 44822). In the reevaluation process, there is no specification (or mandate) to replicate previously used instruments, although this is usually what happens. In the reevaluation of students with SLD, another comprehensive intellectual assessment typically is provided. Given time constraints, the readministration of a comprehensive intellectual measure may not be time- or cost-effective practice, particularly if the test yields relatively unchanged ability estimates. The use of an intellectual screening test to recheck the intellectual status of the referred student might save time that would be better spent evaluating the effectiveness of the individual education program (Ross-Reynolds, 1990) or in the provision of other types of services (e.g., consultation, counseling, research, and program development).

One of the recommended uses of the Kaufman Brief Intelligence Test (K-BIT) is rechecking the intellectual status of an individual who previously has been administered a comprehensive intelligence test (Kaufman & Kaufman, 1990). In addition, the K-BIT also was developed to measure and compare verbal and nonverbal abilities, as is done with the Wechsler scales. "The relationship between the K-BIT and the Wechsler series of scales is intuitive since the verbal/nonverbal split of the brief test closely resembles the Verbal/Performance dichotomy that characterizes Wechsler's popular tests" (Kaufman & Kaufman, 1990, p. 7). Although studies to date indicate a high degree of concurrent and convergent validity when the K-BIT is compared to more comprehensive intellectual ability measures (Kaufman & Kaufman, 1990; Naugle et al., 1993; Prewett, 1992a, 1992b), little is known about the psychometric relations between the K-BIT and the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991). Canivez (1995) and Prewett (1995) have shown high positive correlations between the K-BIT and the WISC-III; however, Canivez (1995) has shown there to be no agreement between Vocabulary-Matrices discrepancies on the K-BIT and VIQ-PIQ discrepancies on the WISC-III. At present, there are no published studies that have investigated the validity of the K-BIT among a sample of students with SLD. The use of the K-BIT in the reevaluation of students with SLD requires careful study to determine whether it yields similar information and diagnostic power when compared to a more comprehensive measure, such as the WISC-III.

The present study investigated the concurrent validity of the K-BIT in comparison with the WISC-III in a sample of students previously identified with SLD. Examination of the diagnostic utility of the K-BIT in identifying the presence or absence of severe achievement-ability discrepancies when compared to the presence or absence of severe achievement-ability discrepancies identified by the WISC-III was of particular importance. If the K-BIT is to be a useful instrument in reducing the reevaluation time of students with SLD by supplanting a comprehensive intellectual ability measure, it must have acceptable levels of diagnostic agreement when compared to results obtained from a comprehensive intellectual ability measure.

METHOD

Subjects

The 75 participants in the present study were elementary (K-6) and middle school (6-8) students in a major Southwestern metropolitan public school system who were referred for triennial multidisciplinary reevaluations. All previously had been identified as specific learning disabled by multidisciplinary evaluation teams (MET) according to state special education rules and regulations recommended for classification of students as learning disabled; these regulations were similar to those specified by the United States Department of Education (1992). Learning disability was defined operationally as a severe discrepancy between ability and achievement using a regression approach (Heath & Kush, 1991; Reynolds, 1984; Wilson & Cone, 1984), and 1.5 standard errors of estimate was suggested as a minimum criterion for severe discrepancy when the students initially had been classified by multidisciplinary evaluation teams. Sixty-five percent of the sample ($n = 49$) were male, 35% ($n = 26$) were female, and the mean age of the students was 11.79 years ($SD = 2.04$, range = 6 to 15 years). Ethnic characteristics of the students were as follows: Caucasian, 35% ($n = 26$); Black, 9% ($n = 7$); Hispanic, 44% ($n = 33$); Native American, 11% ($n = 8$); and Hispanic-Native American, 1% ($n = 1$). All students in this study were sufficiently proficient in English for appropriate administration of present instruments, although some were bilingual. Bilingual students were evaluated by a bilingual school psychologist who determined that their English skills were adequate for proper administration of selected tests. Sixty-nine percent ($n = 52$) were monolingual English speakers, while 17% ($n = 13$) had English as a primary language and Spanish as a secondary language, and 13% ($n = 10$) had Spanish as a primary language and English as a secondary language.

Measures

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 ($N = 2{,}022$) that closely approximated 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; these ranged 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).

Wechsler Intelligence Scale for Children-Third Edition. The Wechsler Intelligence Scale for Children-Third Edition (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 comprised of several subtests that measure different aspects of intelligence and yields three composite IQs (viz., Verbal [VIQ], Performance [PIQ], and Full Scale [FSIQ]), which provide estimates of the individual's verbal, perceptual/nonverbal, and general intellectual abilities. The WISC-III also yields 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) that closely approximated 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 excellent; these ranged from .80 to .97 within the 11 age levels with 55 of 77 (71%) coefficients $\geq$ .90. Average test-retest stability estimates for the three IQ and four Index scores also were excellent and ranged from .82 to .94. Concurrent validity studies generally found moderately high correlations with other intellectual ability measures. VIQ tended to correlate more highly with verbal ability measures than nonverbal ability measures, while PIQ tended to correlate more highly with nonverbal ability measures than with verbal ability measures (Wechsler, 1991) as expected.

Woodcock-Johnson-Revised Tests of Achievement. The Woodcock-Johnson-Revised Tests of Achievement (WJ-R ACH) is an individually administered test of academic achievement that asseses various aspects of reading, mathematics, writing, and general knowledge. Achievement subtests in the Standard Battery (viz., Letter-Word Identification, Passage Comprehension, Calculation, Applied Problems, Dictation, Writing Samples, Science, Social Studies, and Humanities) combine to form four achievement clusters (viz., Broad Reading, Broad Mathematics, Broad Written Language, and Broad Knowledge). Internal consistency reliability coefficients were high, with *Mdn* correlations that ranged from .87 to .93 across the entire age range for the Standard Battery. Concurrent validity data presented in the Examiner's Manual (Woodcock & Mather, 1989) indicated that the WJ-R ACH clusters correlated moderately well with other measures of academic achievement that assess similar domains.

Procedure

Students were administered the K-BIT, WISC-III, and WJ-R ACH as part of comprehensive triennial multidisciplinary reevaluations. The K-BIT and WISC-III were administered in counterbalanced order, during the same test session, by one of three licensed and Nationally Certified School Psychologists. The WJ-R ACH was administered, in most instances, by the student's special education teacher; however, for some students, the WJ-R ACH was administered by the school psychologist. K-BIT Vocabulary, Matrices, and IQ Composite standard scores were obtained, and Vocabulary-Matrices discrepancy scores were evaluated for significant differences at the α = .05 and α = .01 levels. (See Table C.5, Kaufman & Kaufman, 1990, p. 112.) Vocabulary-Matrices discrepancy scores also were evaluated for "abnormality" based upon a 5% population prevalence criterion. (See Table 3.2, Kaufman & Kaufman, 1990, p. 46.)

WISC-III VIQ, PIQ, FSIQ, VCI, POI, FDI, and PSI scores were obtained but of the 75 students, 2 were not administered the Symbol Search subtest. Thus, analyses for the Processing Speed Index are based on $n = 73$. VIQ-PIQ discrepancy scores were evaluated for significance for $\alpha = .05$ (see Table B.1, Wechsler, 1991, p. 261) and $\alpha = .01$. Critical values for VIQ-PIQ significance for $\alpha = .01$ are not available in the *WISC-III Manual* and while Naglieri (1993) provided critical values for significant VIQ-PIQ differences ($\alpha = .01$), these values are inflated for use in this study due to Bonferroni correction, which adjusts for the familywide error rate in multiple discrepany comparisons. The present study examined only one WISC-III pairwise comparison (viz., VIQ - PIQ), so critical values for significance for $\alpha = .01$ were obtained using the following formula:

$$\text{Difference Score} = z\sqrt{SEM_a^2 + SEM_b^2},$$

where $z = 2.5758$ (value from the normal curve corresponding to $\alpha = .01$), 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 *SEM*s used for each age level were obtained from Table 5.2 in the *WISC-III Manual* (Wechsler, 1991, p. 168). VIQ-PIQ discrepancies also were considered "abnormal" at or below the 5% population prevalence criterion level. (See Table B.2, Wechsler, 1991, p. 262.)

Raw scores from the WJ-R ACH were converted to standard scores ($M = 100$, $SD = 15$) based upon age norms found in Tables B and D (Woodcock & Johnson, 1989). Most students were administered all WJ-R ACH subtests, however, some were administered subtests related only to their suspected disabilities.

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 (dependent variable) in a linear regression analysis. Diagnostic efficiency statistics were calculated as recommended by Kessel and Zimmerman (1993) and automated by Canivez and Watkins[1] (in press) to evaluate further the K-BIT Vocabulary-Matrices discrepancy. Kappa (κ) coefficients (Cohen, 1960) were calculated to assess the degree of agreement between Vocabulary-Matrices and VIQ-PIQ discrepancies for $\alpha = .05$, $\alpha = .01$, and for the 5% population prevalence criterion. To test whether kappa coefficients were significant, Z-tests were performed as recommended by Fleiss (1981, p. 219). Pearson product-moment correlation coefficients also were obtained between the K-BIT and WJ-R ACH and between the WISC-III and WJ-R ACH.

The second investigation of the present study was determining the level of agreement between severe achievement-ability discrepancies identified with K-BIT IQ Composite scores and WJ-R ACH subtests with severe achievement-ability discrepancies identified with WISC-III FSIQ scores and WJ-R ACH subtests. Predicted achievement was obtained using the formula: Predicted Achievement =

[1]The standard 2 × 2 diagnostic efficiency table (Canivez & Watkins, in press) was modified to a 3 × 3 table to accommodate the three possibilities of verbal-nonverbal ability discrepancy results (viz., not significant, VIQ/Vocabulary > PIQ/Matrices, or PIQ/Matrices > VIQ/Vocabulary).

$r_{xy}(IQ - M_{IQ}) + M_{ACH}$, where IQ = the obtained IQ score, $M_{IQ} = 100$ (average IQ score), and $M_{ACH} = 100$ (average achievement score). Because the actual IQ-Achievement correlations found in the present study are likely to underestimate the true relationships in the general population due to restricted range and because the relationships between the WISC-III and WJ-R ACH and K-BIT and WJ-R ACH in the general population were not known; $r_{xy} = .65$ (Heath & Kush, 1991). Severe discrepancy between predicted achievement and actual achievement was defined by the formula:

$$D > 15z\sqrt{1 - r_{xy}},$$

(Reynolds, 1984), where D = Predicted Achievement − Actual Achievement, $z = 1.65$ (z corresponds to $\alpha = .05$ in a one-tailed significance test), and $r_{xy} = .65$ (median IQ-Achievement correlation recommended by Heath & Kush, 1991).

Diagnostic efficiency tables (Canivez & Watkins, in press) that compared the presence or absence of severe achievement-ability discrepancies between the K-BIT predicted achievement and WJ-R ACH with the presence or absence of severe achievement-ability discrepancies between the WISC-III predicted achievement and WJ-R ACH were created, and diagnostic efficiency statistics were calculated as recommended by Kessel and Zimmerman (1993). To test whether kappa coefficients were significant, Z-tests were performed as recommended by Fleiss (1981, p. 219).

RESULTS

Concurrent Validity

Pearson product-moment correlation coefficients and r^2s for the K-BIT and WISC-III are presented in Table 1. All correlations except PSI were significant ($p < .0001$). Correlations ranged from .18 to .82 ($M_r = .62$[2]). The magnitude of these results was somewhat surprising given the restricted range that clinical samples normally yield. Judging from the standard deviations and range from the WISC-III and K-BIT, this sample had a restricted range compared to the standardization samples. Consistent with previous investigations between the K-BIT and WISC-R (Kaufman & Kaufman, 1990; Prewett, 1992a, 1992b), WISC-III (Prewett, 1995), and WAIS-R (Naugle et al., 1993), the K-BIT IQ Composite correlated significantly with the WISC-III FSIQ ($r = .82$), and 67% of the variability of FSIQ was accounted for by the K-BIT IQ Composite. Differences between correlation coefficients were tested using Hotelling's formula for a t-test when coefficients of correlation are correlated (Guilford & Fruchter, 1978, p. 164). As expected, the Vocabulary subtest had a significantly higher correlation with the WISC-III VIQ than with PIQ, $t(72) = 3.32$, $p < .001$ and significantly higher correlation with VCI than with POI, $t(72) = 2.89$, $p < .005$. The Matrices subtest correlated equally well with PIQ, POI, VIQ, and VCI; no significant differences were noted among the correlations. Similar correlational results were obtained in other studies (Kaufman & Kaufman, 1990; Prewett, 1992a). The lowest correlations were with the PSI as expected.

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

Table 1
Pearson Product-Moment Correlation Coefficients between the K-BIT and WISC-III (N = 75)

	K-BIT Vocabulary	K-BIT Matrices	K-BIT IQ Composite
WISC-III			
VIQ	.72 (.51)	.60 (.36)	.81 (.66)
PIQ	.51 (.26)	.64 (.41)	.71 (.50)
FSIQ	.67 (.45)	.67 (.45)	.82 (.67)
VCI	.73 (.53)	.56 (.31)	.78 (.61)
POI	.55 (.30)	.62 (.38)	.72 (.52)
FDI	.48 (.23)	.56 (.31)	.64 (.41)
PSI[a]	.18 (.03)*	.26 (.07)**	.27 (.07)**

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. All correlations significant $p < .0001$ except where noted.
r^2s presented in parentheses.
[a]$n = 73$.
**ns*. ***p* < .05.

Descriptive statistics for the K-BIT and WISC-III are presented in Table 2. Bonferroni correction for the five logical mean score comparisons below resulted in an adjusted $\alpha = .01$. Students' K-BIT IQ Composite and WISC-III Full Scale IQ scores did not differ, $t(74) = 1.62$, *ns*, nor did their K-BIT Vocabulary subtest and WISC-III VIQ scores, $t(74) = 1.28$, *ns*. No significant differences existed between students' K-BIT Vocabulary subtest and WISC-III VCI scores, $t(74) = 0.28$, *ns*. However, students obtained significantly lower K-BIT Matrices subtest scores than WISC-III PIQ scores, $t(74) = 2.91$, $p < .005$ and lower K-BIT Matrices scores than WISC-III POI scores, $t(74) = 3.80$, $p < .001$. These results also were

Table 2
Descriptive Statistics for K-BIT and WISC-III Scores (N = 75)

	M	*SD*	Range
WISC-III			
VIQ	78.05	13.08	55 - 113
PIQ	90.65	13.59	62 - 126
FSIQ	82.68	13.07	59 - 119
VCI	79.17	13.39	52 - 108
POI	92.20	14.82	60 - 131
FDI	79.77	10.67	61 - 112
PSI[a]	92.82	12.61	64 - 122
K-BIT			
Vocabulary	79.49	13.16	44 - 102
Matrices	86.65	14.37	56 - 130
IQ Composite	81.27	12.32	57 - 108

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, K-BIT = Kaufman Brief Intelligence Test.
[a]$n = 73$.

found with the WISC-R (Prewett, 1992a, 1992b). Naugle et al. (1993) also reported significant, but small, differences; subjects scored consistently *higher* on the K-BIT.

The regression analysis that assessed the ability of the K-BIT Vocabulary-Matrices discrepancy score to predict the WISC-III VIQ-PIQ discrepancy score was significant, $F(1, 73) = 5.68$, $p < .02$. However, only 7% of the variability in WISC-III VIQ-PIQ discrepancy was accounted for by the K-BIT Vocabulary-Matrices discrepancy. Naugle et al. (1993) found that the K-BIT Vocabulary-Matrices discrepancy accounted for only 21% of the variability in WAIS-R VIQ-PIQ discrepancies. VIQ-PIQ discrepancies ($M = -12.60$, $SD = 10.33$) were also significantly larger than Vocabulary-Matrices discrepancies ($M = -7.16$, $SD = 16.03$), $t(74) = 2.84$, $p < .006$. Table 3 presents frequency data for students who demonstrated various K-BIT Vocabulary-Matrices discrepancies and WISC-III VIQ-PIQ discrepancies for $\alpha = .05$ and .01 and for the 5% population prevalence level. Table 4 presents the diagnostic efficiency statistics for these comparisons.

Table 3

Frequencies of Students Who Showed 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	15_d	0_c	18_c
Vocabulary > Matrices	9_b	0_a	2
Matrices > Vocabulary	11_b	0	20_a
	WISC-III VIQ-PIQ		
α = .01	*ns*	VIQ > PIQ	PIQ > VIQ
K-BIT Vocabulary-Matrices			
ns	30_d	0_c	20_c
Vocabulary > Matrices	6_b	0_a	0
Matrices > Vocabulary	11_b	0	8_a
	WISC-III VIQ-PIQ		
≤5% Population Prevalence	> 5%	VIQ > PIQ	PIQ > VIQ
K-BIT Vocabulary-Matrices			
> 5%	56_d	0_c	10_c
Vocabulary > Matrices	1_b	0_a	0
Matrices > Vocabulary	8_b	0	0_a

Note.—WISC-III = Wechsler Intelligence Scale for Children-Third Edition; VIQ = Verbal IQ; PIQ = Performance IQ; K-BIT = Kaufman Brief Intelligence Test. Numbers along the diagonal 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).

These data indicated that for α = .05 and .01 and for the 5% population prevalence level, kappa coefficients were not significant and represented chance levels of agreement between Vocabulary-Matrices and VIQ-PIQ discrepancies.

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	.53	.29	.00
Specificity	.43	.64	.86
Positive Predictive Power	.50	.32	.00
Negative Predictive Power	.45	.60	.85
False Positive Rate	.57	.36	.14
False Negative Rate	.47	.71	1.00
Overall Correct Classification	.47	.51	.75
Kappa (κ)	.07	−.01	−.14
SE_κ	.09	.10	.11
Z	.77	−.12	−1.23
p	*ns*	*ns*	*ns*

Note.—PP = Population prevalence.

Diagnostic Agreement

Tables 5 and 6 present correlation coefficients between the K-BIT and WJ-R ACH and WISC-III and WJ-R ACH, respectively. Correlations are generally lower (but still significant) than those found between the K-BIT and WISC-III and provide evidence for construct validity because individual intelligence and achievement tests (heterotrait-monomethod) are designed to measure somewhat different domains (Campbell & Fiske, 1959). Verbal ability estimates (Vocabulary and VIQ) correlated as highly or more highly with WJ-R ACH than general intellectual ability estimates (IQ Composite and FSIQ). It is also interesting to note that nonverbal or perceptual ability measures (Matrices and PIQ) yielded lower correlations with WJ-R ACH than verbal ability measures (Vocabulary and VIQ) and general

Table 5
Pearson Product-Moment Correlation Coefficients between the K-BIT and WJ-R ACH

		K-BIT		
	n	Vocabulary	Matrices	IQ Composite
WJ-R ACH				
LWID	67	.48 (.23)****	.17 (.03)	.39 (.15)***
PC	67	.58 (.34)****	.36 (.13)**	.57 (.32)****
C	66	.45 (.20)***	.25 (.06)*	.42 (.18)***
AP	66	.50 (.25)****	.43 (.18)***	.56 (.31)****
D	70	.45 (.20)****	.12 (.01)	.33 (.11)**
WS	74	.44 (.19)****	.28 (.08)*	.44 (.19)****
BR	68	.57 (.32)****	.26 (.07)*	.50 (.25)****
BM	66	.55 (.30)****	.40 (.16)***	.58 (.34)****
BWL	69	.54 (.29)****	.29 (.08)*	.50 (.25)****

Note.—K-BIT = Kaufman Brief Intelligence Test; WJ-R ACH = Woodcock-Johnson-Revised Tests of Achievement; LWID = Letter-Word Identification; PC = Passage Comprehension; C = Calculation; AP = Applied Problems; D = Dictation; WS = Writing Samples; BR = Broad Reading; BM = Broad Mathematics; BWL = Broad Written Language.
r^2s presented in parentheses.
*$p < .05$. **$p < .01$. ***$p < .001$. ****$p < .0001$.

Table 6
Pearson Product-Moment Correlation Coefficients between the WISC-III and WJ-R

		WISC-III		
	n	VIQ	PIQ	FSIQ
WJ-R ACH				
LWID	67	.30 (.09)**	.08 (.01)	.21 (.04)
PC	67	.57 (.32)****	.39 (.15)***	.53 (.28)****
C	66	.42 (.18)***	.29 (.08)*	.38 (.14)***
AP	66	.65 (.42)****	.46 (.21)****	.61 (.37)****
D	70	.28 (.08)*	.11 (.01)	.21 (.04)
WS	74	.58 (.34)****	.44 (.19)****	.56 (.31)****
BR	68	.46 (.21)****	.23 (.05)*	.38 (.14)***
BM	66	.63 (.40)****	.44 (.19)***	.59 (.35)****
BWL	69	.52 (.27)****	.35 (.12)**	.47 (.22)****

Note.—WISC-III = Wechsler Intelligence Scale for Children-Third Edition; VIQ = Verbal IQ; PIQ = Performance IQ; FSIQ = Full Scale IQ; WJ-R ACH = Woodcock-Johnson-Revised Tests of Achievement; LWID = Letter-Word Identification; PC = Passage Comprehension; C = Calculation; AP = Applied Problems; D = Dictation; WS = Writing Samples; BR = Broad Reading; BM = Broad Mathematics; BWL = Broad Written Language. r^2s presented in parentheses.
*$p < .05$. **$p < .01$. ***$p < .001$. ****$p < .0001$.

intellectual ability estimates (IQ Composite and FSIQ) as expected (Kaufman & Kaufman, 1990; Wechsler, 1991).

Table 7 presents descriptive statistics for the WJ-R ACH, and examination of mean scores indicated that this sample of students previously identified as learning disabled as a group scored approximately one standard deviation below the mean (except for Applied Problems) for the standardization sample. Because of the heterogeneous nature of groups of "learning disabled" students (learning disability may exist in any *one* of seven areas), clinical group mean scores may appear higher than expected when compared to mean WJ-R ACH scores for the standardization sample.

Table 7
Descriptive Statistics for WJ-R Tests of Achievement

WJ-R ACH	*n*	*M*	*SD*	Range
Letter-Word Identification	67	79.96	10.41	53 - 101
Passage Comprehension	67	87.06	12.22	48 - 131
Calculation	66	77.02	10.41	53 - 102
Applied Problems	66	91.12	11.93	60 - 132
Dictation	70	71.04	10.31	32 - 88
Writing Samples	74	85.82	16.81	32 - 126
Broad Reading	68	81.25	11.27	52 - 113
Broad Mathematics	66	81.29	11.64	54 - 119
Broad Written Language	69	75.30	10.73	31 - 97

Note.—WJ-R ACH = Woodcock-Johnson-Revised Tests of Achievement.

Table 8 presents diagnostic efficiency statistics for each of the WJ-R ACH subtests and global achievement scores and shows that kappa coefficients ranged from .31

to 1.0 (Mdn_κ = .65[3]). All kappa coefficients were significant and indicated that the agreement for the presence or absence of severe achievement-ability discrepancies between the K-BIT and WJ-R ACH with the presence or absence of severe achievement-ability discrepancies between the WISC-III and WJ-R ACH was well beyond chance. In fact, 78% (7 of 9) of the kappa coefficients were in the substantial or almost perfect agreement range (Everitt & Hay, 1992). One kappa coefficient indicated perfect agreement between the K-BIT and WISC-III in identifying presence or absence of severe discrepancy for the Applied Problems subtest. Other indices of diagnostic efficiency also yielded encouraging, positive results. Agreement between the K-BIT and WISC-III in identifying severe discrepancies was reflected in high levels of positive predictive power, negative predictive power, and overall correct classification (hit rate). Positive predictive power referred to the proportion of students with severe achievement-ability discrepancies identified by the K-BIT who truly showed severe achievement-ability discrepancies with the WISC-III. Negative predictive power was indicated by the proportion of students who did not show severe achievement-ability discrepancies with the K-BIT who also did not show severe achievement-ability discrepancies with the WISC-III. Overall correct classification is the proportion of students correctly classified with and without severe achievement-ability discrepancies with the K-BIT. Generally low false positive rates also were observed. However, in some cases the false negative rate was moderately high.

Table 8
Diagnostic Efficiency Statistics of Severe WJ-R ACH-K-BIT and WJ-R ACH-WISC-III Discrepancies

	LWID	PC	C	AP	D	WS	BR	BM	BWL
Sensitivity	.58	.50	.79	1.00	.84	.67	.40	.40	.67
Specificity	.96	1.00	.94	1.00	.85	.97	.98	.93	.94
Positive predictive power	.78	1.00	.79	1.00	.81	.75	.80	.33	.80
Negative predictive power	.91	.97	.94	1.00	.87	.95	.90	.95	.89
False positive rate	.04	.00	.06	.00	.15	.03	.02	.07	.06
False negative rate	.42	.50	.21	.00	.16	.33	.60	.60	.33
OCC rate	.90	.97	.91	1.00	.84	.93	.90	.89	.87
Kappa (κ)	.61	.65	.73	1.00	.68	.67	.48	.31	.64
SE_κ	.12	.11	.12	.13	.12	.12	.11	.12	.12
Z	5.03	5.71	5.92	7.58	5.71	5.75	4.29	2.50	5.37

Note.—WJ-R ACH = Woodcock-Johnson-Revised Tests of Achievement; WISC-III = Wechsler Intelligence Scale for Children-Third Edition; K-BIT = Kaufman Brief Intelligence Test; LWID = Letter-Word Identification; PC = Passage Comprehension; C = Calculation; AP = Applied Problems; D = Dictation; WS = Writing Samples; BR = Broad Reading; BM = Broad Mathematics; BWL = Broad Written Language; OCC = Overall Correct Classification.
All kappa coefficients significant $p < .0001$ except for Broad Mathematics $p < .01$.

DISCUSSION

The K-BIT IQ Composite, Vocabulary, and Matrices scores compared favorably to the WISC-III IQs and Index scores, and these data provided ample evidence

[3]Although Kessel and Zimmerman (1993) recommend that frequency data be reported routinely, space limitations prohibited presentation of these data. Diagnostic efficiency tables complete with frequency data are presented in Canivez (1996).

in support of concurrent validity of the K-BIT as a valid brief estimate of general intellectual abilities in reevaluation of elementary and middle-school children with learning disability.

Partly due to restricted range, the K-BIT Vocabulary, Matrices, and IQ Composite and WISC-III VIQ, PIQ, FSIQ, VCI, and POI correlations obtained in the present study were lower than correlations between WISC-III Vocabulary and Block Design with VIQ, PIQ, FSIQ, VCI, and POI presented in the *WISC-III Manual*. (See p. 281, Wechsler, 1991.) These two subtests (Vocabulary and Block Design) frequently are combined in a two-subtest short form for intellectual screening purposes (Kaufman, 1990; Sattler, 1992). Silverstein (1990) has argued, however, 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 gathered from administrations in which the students were administered the entire test, and the resulting scores may not correspond if only the short-form subtests were administered in isolation (Silverstein, 1990). While the K-BIT retains high correlations with the various WISC-III IQ and Index scores, it has motor-free subtests, which is an advantage with various students who have physical disabilities.

The present study indicated that the K-BIT falls short of its goal of assessing a verbal-nonverbal dichotomy in that there was 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 (7%), 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 more comprehensive intelligence tests, such as the WISC-III (or WAIS-R, Naugle et al., 1993). These data suggested that the K-BIT does not possess adequate sensitivity or positive predictive power to identify correctly students who have VIQ-PIQ discrepancies. This may be related in part to the fact that the K-BIT is comprised of only two subtests and does not sample the respective domains as well as a more comprehensive intellectual measure. Kaufman and Kaufman noted that "interpretations of standard score differences between the K-BIT subtests, and their possible relationship to scores on other tests such as Wechsler's Verbal IQ/Performance IQ difference, are inferential and hypothetical" and "One cannot generalize about an entire construct from a single subtest" (1990, p. 43). It also may be due to the unreliability of difference (discrepancy) scores (Silverstein, 1981; Thorndike & Hagen, 1977).

Another possibility for this low agreement is that some argue that the WISC-III may reflect only a measure of general intelligence (*g*) rather than a Verbal-Performance model, where "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). Macmann and Barnett (1994) also suggested that nonverbal indices like the PIQ and POI may only be less reliable measures of general intelligence than VIQ and VCI. Factor-structure matrices presented by Macmann and Barnett show that WISC-III verbal subtests (Vocabulary, Information, Similarities, and Comprehension) loaded as well for the

Performance Factor as some performance subtests (Picture Completion and Picture Arrangement). This also could help to explain why, in the present study and in other research (Kaufman & Kaufman, 1990), the K-BIT Vocabulary subtest correlated more highly with VIQ and VCI than PIQ and POI, but the Matrices subtest correlated equally well with the WISC-III VIQ, VCI, PIQ, and POI. Kaufman and Kaufman (1990) and Prewett (1992a) offered no insights with regard to equivalent correlations between Matrices and VIQ or PIQ. In the present study, VIQ and PIQ were correlated moderately (r = .67), as were VCI and POI (r = .64). Interestingly, the correlation between Vocabulary and Matrices (r = .30) in the present study suggested significantly less overlap than VIQ (VCI) and PIQ (POI). This, however, also may reflect greater error variance due in part to substantially shorter scales. Alternatively, verbal-nonverbal differences may not have been in agreement because the Matrices subtest was designed to be a measure of fluid (Gf) abilities, whereas the PIQ may be thought of as reflecting Horn's *Visual General Ability* factor (Gv) rather than fluid (Gf) abilities (Carroll, 1993; Sattler, 1992; Woodcock, 1990).

The comparison of Vocabulary and Matrices subtest scores in predicting or hypothesizing VIQ-PIQ differences appears to be a questionable practice based on these data because it did not provide significant insight into possible verbal and nonverbal differences on the WISC-III. Kaufman and Kaufman (1990) were justifiably cautious in recommending that no inferences be made about possible verbal and nonverbal differences with the K-BIT. However, they provide no empirical support for the "mandate" (p. 46) for recommending administration of a comprehensive intellectual battery to investigate abnormal Vocabulary-Matrices discrepancies. The present study suggests that this "mandate" may not be justified in regard to the WISC-III given the low positive predictive power; low sensitivity; and low, nonsignificant kappa coefficients. At the present time, the K-BIT should be considered only as an estimate of general intelligence (g) until additional research can help to investigate further the nature of and relationships between the Vocabulary and Matrices subtests with other instruments and populations hypothesized to reflect multiple intellectual factors.

Regardless of advantages, use of brief intellectual measures (K-BIT or short forms) has not been recommended for making educational or diagnostic decisions (Kaufman, 1990; Sattler, 1992; Silverstein, 1990). Kaufman (1990) and Sattler (1992) 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 little to no empirical support for profile or ipsative analysis in *comprehensive* intellectual ability measures (Hale, 1979; Hale & Landino, 1981; Hale & Saxe, 1983; McDermott et al., 1990; McDermott et al., 1992; McDermott et al., 1989; McDermott, Glutting, Jones, Watkins, & Kush, 1989; Watkins & Kush, 1994).

The diagnostic utility of the K-BIT was addressed in the present study, and the K-BIT was found to be extremely useful in that achievement-ability discrepancies found between the K-BIT and WJ-R ACH had very high positive predictive power, negative predictive power, and overall correct classification when compared to achievement-ability discrepancies found between the WISC-III and WJ-R ACH. False positive classifications were quite low and were indicated by students who showed severe achievement-ability discrepancies by the K-BIT, but did not show the corresponding severe achievement-ability discrepancies with the WISC-III. False negative rates were, in some cases, moderately high. False negative classifications were indicated by those who showed no severe achievement-ability discrepancies with the K-BIT, but showed severe achievement-ability discrepancies with the WISC-III. In a reevaluation situation, a false positive classification would result in continuing to classify a student's achievement as "discrepant" when it "truly" is not (based on a more comprehensive intellectual ability measure). This result might lead a MET to continue eligibility for special education for that student when it only used a brief intellectual measure. A false negative classification would result in determining that the student's achievement is not "discrepant" when it "truly" was discrepant (based on a comprehensive intellectual ability measure). This result might lead a MET to terminate special education programming eligibility when only using a brief intellectual measure. Both of these situations are possible in any assessment that uses a brief *or* comprehensive intellectual measure due to measurement error, and there is yet no agreement as to which is the more serious error (Heath & Kush, 1991; Reynolds, 1984).

If the present results are replicated, then the K-BIT may supplant a comprehensive intellectual ability measure in reevaluations of students with SLD while retaining a high degree of diagnostic agreement. This practice could save considerable time in the reevaluation process that could be better spent in alternative assessment practices or in providing alternative services, such as program development, counseling, or research. Future research should continue to examine the relationship of the K-BIT with other comprehensive intellectual ability measures and with different samples of normal individuals and clinical groups in order to further define and delineate its psychometric characteristics. Differences between racial or ethnic groups, as well as in bilingual students, should also be explored. As with comprehensive intellectual ability measures, it will be important to determine whether there is differential validity for different subgroups in the population.

REFERENCES

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

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. (1996). 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 No. ED385959)

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

Carroll, J. B. (1993). *Human cognitive abilities of factor-analytic studies.* Cambridge, England: Cambridge University Press.

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

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: John Wiley.

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

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

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.

Heath, C. P., & Kush, J. C. (1991). Use of discrepancy formulas in the assessment of learning disabilities. In J. E. Obrzut and G. W. Hynd (Eds.), *Neuropsychological foundations of learning disabilities: A handbook of issues, methods, and practice* (pp. 287-307). San Diego, CA: Academic Press.

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.

Hutton, J. B., Dubes, R., & Muir, S. (1992). Assessment practices of school psychologists: Ten years later. *School Psychology Review, 21*, 271-284.

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

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. (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.

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.

Reschley, D. J., Genshaft, J., & Binder, M. S. (1987). *The 1986 NASP survey: Comparison of practitioners, NASP leadership, and university faculty on key issues*. Washington: National Association of School Psychologists.

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

Ross, R. P. (1992). Accuracy in analysis of discrepancy scores: A nationwide study of school psychologists. *School Psychology Review, 21*, 480-493.

Ross-Reynolds, J. (1990). Best practices in conducting reevaluations. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-II* (pp. 195-206). Washington: National Association of School Psychologists.

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.

Smith, D. K. (1984). Practicing school psychologists: Their characteristics, activities, and populations served. *Professional Psychology: Research and Practice, 15*, 798-810.

Thorndike, R. L., & Hagen, E. P. (1977). *Measurement and evaluation in psychology and education* (4th ed.). New York: John 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 way? *School Psychology Review, 23*, 638-649.

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

Wilson, L. R., & Cone, T. (1984). The regression equation method of determining academic discrepancy. *Journal of School Psychology, 22*, 95-110.

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

Woodcock, R. W., & Johnson, M. B. (1989). WJ-R Tests of Achievement: Norm Tables. In R. W. Woodcock & M. B. Johnson, *Woodcock-Johnson Psycho-Educational Battery-Revised.* Allen, TX: DLM Teaching Resources.

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.