1-1-2005

Comparison of the Kaufman brief intelligence test-second edition and the Wechsler intelligence scale for children-fourth edition

Ryan L. Mitchell
Eastern Illinois University

THESIS REPRODUCTION CERTIFICATE

TO: Graduate Degree Candidates (who have written formal theses)

SUBJECT: Permission to Reproduce Theses

The University Library is receiving a number of request from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow these to be copied.

PLEASE SIGN ONE OF THE FOLLOWING STATEMENTS:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

Ryan Mitchell _____ 8-8-2005

Author's Signature Date

I respectfully request Booth Library of Eastern Illinois University **NOT** allow my thesis to be reproduced because:

_____

_____

_____

_____ _____

Author's Signature Date

This form must be submitted in duplicate.

Comparison of the Kaufman Brief Intelligence Test-Second Edition and the Wechsler Intelligence Scale for Children-Fourth Edition

(TITLE)

BY

Ryan L. Mitchell

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Specialist in School Psychology

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

2005
YEAR

I HEREBY RECOMMEND THAT THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

8/8/05
DATE

THESIS DIRECTOR

8/8/05
DATE

DEPARTMENT/SCHOOL HEAD

Running head: Comparison of the Kaufman Brief Intelligence Test-Second Edition

Comparison of the Kaufman Brief Intelligence Test-Second Edition and the Wechsler Intelligence Scale for Children-Fourth Edition

Ryan L. Mitchell

Eastern Illinois University

Abstract

Past research on the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990) and the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) have shown good reliability and validity between the scales. However, recently, the KBIT-2 and the WISC-IV have been published. The purpose of the current study was to independently evaluate the construct validity of the KBIT-2 with the WISC-IV. The participants in this study consisted of 48 students enrolled in kindergarten through ninth grade from school districts in central Illinois, who were referred for initial or triennial multidisciplinary evaluation. The results were as hypothesized, with high statistically significant correlations between the KBIT-2 and WISC-IV composite scores, supporting convergent validity. However, the mean KBIT-2 IQ Composite was significantly higher than the WISC-IV FSIQ and represented a medium effect size. Furthermore, the mean KBIT-2 VIQ was significantly higher than the WISC-IV VCI and showed a moderate effect size. However, the mean KBIT-2 NIQ was not significantly higher than the WISC-IV PRI. This study supports the use of the KBIT-2 in reevaluations or other situations where a brief measure of cognitive ability can be utilized. Limitations to this study included a small sample size and the inclusion of only Caucasian and African American participants.

Comparison of the Kaufman Brief Intelligence Test-Second Edition and the Wechsler Intelligence Scale for Children-Fourth Edition

The main goal of intelligence tests in the early 1900s was to identify children who had intellectual deficiencies. During this period, Binet and Simon published an intelligence test that was aimed at identifying children who would not benefit from regular education (Wechsler, 2003).

Almost 40 years after the introduction of the first intelligence test, David Wechsler introduced the first version of his intelligence test, the Wechsler-Bellevue Intelligence Scale (Wechsler, 1939). Wechsler based his test on the premise that intelligence was a global entity comprised of specific abilities because it is composed of elements or abilities that are distinct from each other. Based on these ideas, Wechsler selected subtests that measured cognitive aspects of intelligence that he thought were important. These aspects included: verbal comprehension, abstract reasoning, perceptual organization, quantitative reasoning, memory, and speed of performance (WISC-IV; Wechsler, 2003). Wechsler then selected subtests based on these six cognitive aspects. The original Wechsler-Bellevue consisted of 11 subtests, including Information, Arithmetic, Similarities, Vocabulary, Digit Span, Comprehension, Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding. From these 11 subtests, 3 composite scores were calculated. The three composite scores were Verbal IQ (VIQ), Performance IQ (PIQ), and the Full Scale IQ (FSIQ).

The Wechsler-Bellevue was used for 10 years, before a downward extension, the Wechsler Intelligence Scale for Children (WISC), was published in 1949. This test was very similar to the original Wechsler-Bellevue; however, the WISC incorporated the Mazes subtest, which brought the subtest total to 12.

In 1974, Wechsler once again revised and restandardized his test by producing the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974). This test retained the same 12 subtests and three composite scores as the WISC, but the age range moved from 5-15 years to 6-16 years.

The WISC-R was utilized for 17 years, until The Psychological Corporation updated the test and produced the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991). In this revision, the test retained the 12 subtests of the WISC-R but added Symbol Search in an attempt to strengthen the Freedom from Distractibility factor. Instead of Symbol Search strengthening this index, it actually pulled Coding away from Freedom from Distractibility and formed a new factor that was called Processing Speed. The WISC-III retained the three composite IQ scores, but also introduced four new factor-based index scores to represent psychometrically pure domains of cognitive functioning. These scores included the Verbal Comprehension Index score (VCI), the Perceptual Organization Index score (POI), the Freedom from Distractibility Index score (FDI), and the Processing Speed Index score (PSI).

Past research has shown the popularity of using the Wechsler scales to measure children's cognitive ability. Stinnett, Havey, and Oehler-Stinnett (1994) reported that 91% of practicing school psychologists surveyed used the WISC-R and 70% used the WISC-III to evaluate children. Furthermore, these instruments had the highest mean number of reported administrations and the highest importance ratings of the intellectual instruments used. The popularity of the Wechsler scales could be attributed to their favorable evaluations in the literature as well as their acceptable psychometric qualities (Stinnett et al, 1994).

Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV)

Recently, The Psychological Corporation introduced the latest revision of this popular intelligence test: the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003). In this latest revision of the WISC several changes have been noted. Three subtests (Mazes, Picture Arrangement, and Object Assembly) were eliminated from the WISC-III, and four new subtests (Picture Concepts, Letter-Number Sequencing, Matrix Reasoning, and Cancellation) were added. With the elimination of the Mazes subtest and the introduction of these four new subtests, the subtest total is now 15. However, of these 15 subtests, 10 are designated as core subtests used to estimate the Full Scale IQ, while the other 5 are designated as supplemental subtests.

> "The core subtests are administered when composite scores are desired. Supplemental subtests extend the range of cognitive skills sampled, provide additional clinical information, and enable the practitioner to complete additional discrepancy analyses. When necessary, supplemental subtests can also be used as substitutes for core subtests" (Wechsler, 2003, p. 6).

The second major change from the WISC-III to the WISC-IV was the type of scores that are calculated. The WISC-IV retained the Full Scale IQ (FSIQ) score but eliminated the Verbal IQ (VIQ) score and the Performance IQ (PIQ) score. Instead of the VIQ and PIQ scores, the WISC-IV offers the two factor-based index scores related to verbal and perceptual functioning. These index scores include the Verbal Comprehension Index (VCI) score and the Perceptual Reasoning Index (PRI) score, formerly known as the Perceptual Organization Index (POI) score in the WISC-III. The VCI score is based on three core subtests: Similarities, Vocabulary, and Comprehension and two supplemental subtests: Information and Word Reasoning. The PRI

score is based on three core subtests: Block Design, Picture Concepts, and Matrix Reasoning, and one supplemental subtest: Picture Completion. The Processing Speed Index score was retained from the WISC-III but the Freedom from Distractibility Index score was renamed the Working Memory Index (WMI). The WMI score is based on two core subtests: Digit Span and Letter-Number Sequencing, and one supplemental subtest: Arithmetic. The PSI score is based on two core subtests: Coding and Symbol Search, and one supplemental subtest: Cancellation. For each of these indexes, the core subtests are used to calculate the index score, and the supplemental subtests are used in cases where one of the core subtests is spoiled. The core subtests from the VCI, PRI, WMI, and PSI scores make up the FSIQ score.

The standardization sample for the WISC-IV consisted of 2,200 children ages 6:0-16:11. The sample included representative proportions of children according to selected demographic variables. Stratification variables included age, sex, race/ethnicity, parent education level, and geographic region. The number of children representing these variables was based on the U.S. population according to the March 2000 Census data (WISC-IV; Wechsler, 2003).

The reliability of WISC-IV scores was estimated using several different methods. Internal consistency reliability was computed using the split-half method for all subtests except Cancellation, Coding, and Symbol Search because these subtests are timed and the split-half method was not appropriate. Therefore, the test-retest stability coefficient was used as the reliability estimate for these three subtests. The stability coefficients ranged from .79 on the Cancellation and Symbol Search subtests to .90 on the Letter-Number Sequencing subtest. Furthermore, the internal consistency reliability coefficients for the Index scores ranged from .88 on the PSI to .94 on the VCI. The internal consistency for the FSIQ was computed to be .97 (WISC-IV; Wechsler, 2003). .

The short term test-retest reliability coefficients were computed for all subtests, index scores and the FSIQ. The test-retest intervals ranged from 13 to 63 days with a mean interval of 32 days. The test-retest reliability coefficients ranged from .76 (Picture Concepts) to .92 (Vocabulary) for subtests. The test-retest reliability coefficients for the index scores ranged from .86 (PSI) to .93 (VCI) and was .93 for the FSIQ (WISC-IV; Wechsler, 2003).

Validity investigations of the WISC-IV were made with comparisons to the WISC-III, Wechsler Preschool and Primary Scale of Intelligence – Third edition (WPPSI-III, Wechsler, 2002), Wechsler Adult Intelligence Scale – Third edition (WAIS-III, Wechsler, 1997), and the Wechsler Abbreviated Scale of Intelligence (WASI, Wechsler, 1999). The WISC-IV and the WISC-III were administered to 244 children aged 6-16 in counterbalanced order with a mean test-retest interval of 28 days (range 5–67 days). The correlations for the subtests ranged from .62 (Comprehension) to .83 (Information). The correlations for the index scores were moderate to high. The correlations ranged from .72 (WMI-FDI) to .87 (VCI-VIQ). Furthermore, the correlations for composite scores were moderate to high for VCI-VCI (r = .88) and PRI-POI (r = .72). Also, the correlation between the WISC-III FSIQ and the WISC-IV FSIQ was .89 (WISC-IV; Wechsler, 2003).

The WISC-IV and the WPPSI-III were administered to 182 children aged 6-7 in counterbalanced order, with a mean test-retest interval of 22 days (range 9-62 days). The correlations for the subtests ranged from .44 (Picture Concepts) to .74 (Information). The index score correlations between the WISC-IV and the WPPSI-III were also moderate to high. The correlations ranged from .65 (PSI-PSQ) to .83 (VCI-VIQ). Furthermore, the validity coefficient for the FSIQ between the WISC-IV and the WPPSI-III was .89 (WISC-IV; Wechsler, 2003).

The WISC-IV and WAIS-III were administered to 198 children aged 16, in counterbalanced order with a mean test-retest interval of 22 days (range 10-67 days). The correlations for the subtests ranged from .59 (Matrix Reasoning) to .83 (Vocabulary). The correlations for the index scores ranged from .76 (PRI-PIQ) to .86 (VCI-VIQ). Furthermore, the composite scores were moderate to high. The correlation for VCI-VCI was .85 and the correlation for PRI-POI was .73. The correlation for the FSIQ was shown to be high (r = .89) (WISC-IV; Wechsler, 2003).

The WISC-IV and the WASI were administered to 260 children aged 6-16, in counterbalanced order, with a mean test-retest interval of 29 days (range 12-64 days). The correlations for the subtests ranged from .71 (Matrix Reasoning) to .79 (Vocabulary). Furthermore, the WISC-IV FSIQ correlated highly with the WASI FSIQ-4 (.86) and WASI FSIQ-2 (.83) (WISC-IV; Wechsler, 2003).

To further examine the validity of WISC-IV scores, an exploratory factor analysis was used. The initial exploratory factor analysis was conducted using all 2,200 participants and focused on only the core subtests. The overall analysis showed that there was a four-factor structure that consisted of Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed. When the 2,200 individuals were subdivided into four age groups (ages 6-7, 8-10, 11-13, and 14-16) the four-factor structure was also replicated. However, in ages 6-7, the Picture Concepts subtest loaded equally on the Verbal Comprehension factor (.21) and the Perceptual Reasoning factor (.20). Neither loading was salient for Picture Completion for this age group (WISC-IV; Wechsler, 2003).

The second exploratory factor analysis included both core subtests *and* supplementary subtests. This analysis included 1,525 children from the normative sample who also completed

Arithmetic. This factor analysis also showed the same four-factor structure that was observed with the core subtest for all age groups. However, Information had a small secondary loading on the Working Memory factor in ages 6-7 and 8-10 (WISC-IV; Wechsler, 2003).

To further evaluate the four-factor structure, a cross-validation analysis was conducted by randomly selecting seven samples of 440 children. The results showed the stability of the VCI, PRI, WMI, and PSI factors across the random sub-samples and supported the four-factor structure (WISC-IV; Wechsler, 2003).

The confirmatory factor analysis for the core subtests showed that the four-factor structure was the best fit for all age groups. Furthermore, when taking into account the core subtests *and* the supplementary subtests, the four factor structure fit the data best for all age groups; and a five-factor model did not indicate a substantive improvement over the four-factor model (WISC-IV; Wechsler, 2003).

From examining the reliability and validity studies published in the WISC-IV manual, it was clear that there is evidence for the reliability and validity of the WISC-IV scores. This may help to explain why many professionals in the area of assessment continue to choose the WISC-IV for comprehensive intellectual assessment.

Although the Wechsler scales are widely used and effective instrument for evaluating an individual's cognitive ability, they are time consuming to administer. On average the WISC-IV takes between 1 and 2 hours to administer and score. Therefore, a reliable and valid brief measure of intelligence, when appropriate, would be helpful to practioners in order to save time. Also, brief intelligence measures would be helpful for screening and research purposes.

In order to reduce the amount of time it takes to estimate intellectual abilities, some clinicians have used short forms of comprehensive intelligence tests. Short forms are created

using the standardization data where the individuals were 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 (Silverstein, 1990). Kaufman and Kaufman (2001) argued that the development and standardization of well-normed, reliable, and valid brief tests of intelligence would eliminate the need to create short forms of more comprehensive intelligence.

In the past several years, brief intelligence tests that exhibit good reliability and validity have been introduced. One of these tests was the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990).

Research conducted by Prewett and McCaffery (1993) showed the K-BIT would be a good screening instrument when the Stanford-Binet was the follow-up comprehensive measure of intelligence. This was evidenced by a .81 correlation between the Stanford-Binet Intelligence Test Composite and the K-BIT IQ Composite. However, this study also showed that the mean Stanford-Binet Intelligence Test Composite was 5.1 points higher than the mean K-BIT IQ Composite, which was a significant difference and of a moderate effect size.

Hays, Reas, and Shaw (2002) compared the K-BIT to the WASI, which are both brief intelligence tests. They reported that the K-BIT mean score was two points higher than the mean WASI score and that the two tests correlated at .89. Since both of these brief measures are standardized on representative normative samples of the U.S. population, these data suggested that either of these tests would make a good screening instrument.

Several studies have compared the K-BIT to the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1976). Eisenstein and Engelhart (1997) compared the K-BIT to a WAIS-R short form and Axelrod and Naugle (1998) compared the K-BIT IQ Composite to the WAIS-R FSIQ. Eisenstein and Engelhart (1997) reported that the mean K-BIT IQ Composite

was 2.69 points higher than the mean WAIS-R short form score, and they reported a .73 correlation between the two scores. They concluded that the K-BIT adequately estimated the global IQ as measured by the WAIS-R. Axelrod and Naugle (1998) reported similar results in their study. They reported the mean K-BIT IQ Composite score to be 4.3 points higher than the mean WAIS-R FSIQ. Furthermore, they also found a strong correlation between the two test scores ($r = .88$).

There have also been numerous studies comparing the K-BIT to the Wechsler Intelligence Scale for Children-Revised. Prewett (1992a) found the K-BIT IQ Composite score correlated .81 with the WISC-R FSIQ. He also found that the mean WISC-R FSIQ averaged 6.2 points higher than the mean IQ Composite score on the K-BIT. In another study by Prewett (1992b) on 40 incarcerated male juvenile delinquents, the K-BIT IQ Composite score did not significantly differ from the WISC-R FSIQ. Although the correlation in this study was lower than previous studies ($r = .64$), this finding was believed to be due in part to a restriction in range. Although the results of these two studies were not entirely consistent, both studies supported the use of the K-BIT as a screening measure if it is followed by the WISC-R.

Two additional studies were conducted comparing the K-BIT and the WISC-R. Slate and Graham found that the mean WISC-R FSIQ was 1.8 points lower than the mean K-BIT IQ Composite, which was not statistically significant, and they found a correlation between the two IQ scores of .80. Furthermore, Webber and McGillivray (1998) found that the mean WISC-R FSIQ was 2.43 points lower than the mean K-BIT IQ Composite and the two scores correlated .93. Both of these studies showed similar results, as there was a small effect size between score differences.

Once the WISC-III was published, several researchers compared the K-BIT Composite to the WISC-III FSIQ. These studies found that the mean K-BIT IQ Composite score was higher than the mean WISC-III FSIQ (Canivez, 1995; Canivez, 2005; Chin, Ledesma, Cirino, & Sevcik, 2001; Donders, 1995; Grados & Russo-Garcia, 1999; Prewett, 1995; & Thompson, Browne, Schmidt, & Boer, 1997). The difference between the mean scores ranged from .73 (Canivez, 1995) to 8.4 (Donders, 1995) points. Further examination of these studies showed that sample size could be a factor that may explain the differences in the outcomes as smaller samples would have greater sampling error. Whereas, Canivez (1995) had 137 participants in his study, Donders (1995) only had 47 participants in his study. Canivez (2005) obtained a sample size of 207 public school children and adolescents and found that the mean difference between the K-BIT IQ Composite and WISC-III FSIQ was 2.20, once again showing that a larger sample size resulted in a lower mean difference on the composite scores.

The correlations between the WISC-III FSIQ and the K-BIT IQ Composite score were found to be moderate to high. The lowest correlation was reported by Chin et al (2001) with a correlation of .63 between the WISC-III FSIQ and the K-BIT IQ Composite. However, Canivez (1995) reported a much higher correlation of .87 between the K-BIT IQ Composite and the WISC-III FSIQ. Furthermore, Canivez (2005) reported a correlation of .89 between the K-BIT IQ Composite and the WISC-III FSIQ.

Canivez (1996) reported that the K-BIT Vocabulary subtest had a significantly higher correlation with the WISC-III VIQ than with PIQ. Furthermore, there was not a significant difference between the K-BIT Vocabulary subtest score and the WISC-III VIQ score. Also, he reported the Matrices subtest correlated equally with the WISC-III PIQ, POI, VIQ, and VCI. Similar results were reported by Prewett (1992a) using the WISC-R.

Canivez (1995) further reported that the WISC-III VIQ-PIQ discrepancies were significantly larger than Vocabulary-Matrices discrepancies. Prewett (1992b) reported similar results with the WISC-R. He reported that the WISC-R Performance score was consistently higher than the Verbal score, whereas no significant difference was reported between the K-BIT Vocabulary and Matrices subtest. Canivez (1995, 2005) reported that the K-BIT Vocabulary-Matrices discrepancy appeared to be an inadequate predictor of the WISC-III VIQ-PIQ discrepancy, particularly if applied to individuals.

Overall, K-BIT scores have shown acceptable reliability and validity when compared to the WISC. Canivez (1996) reported, "If the past results are replicated, the K-BIT may supplant a comprehensive intellectual ability measure in reevaluations of students with SLD while retaining a high degree of diagnostic agreement." (p. 17) He further, reported that by using a brief measure, one would have more time for other activities such as alternative assessments or providing alternative services.

Kaufman Brief Intelligence Test-Second Edition (KBIT-2)

The KBIT-2 is the first revision of the K-BIT. The KBIT-2 is a brief intelligence test for individuals ages 4-90. The test takes approximately 15-30 minutes to administer and provides Verbal, Nonverbal, and IQ Composite scores.

The Verbal score is comprised of two subtests, Verbal Knowledge and Riddles, which measure crystallized/verbal ability. The Verbal subtests measure verbal and school-related skills by assessing a person's word knowledge, range of general information, verbal concept formation, and reasoning ability (Kaufman & Kaufman, 2004). The Verbal Knowledge subtest is a 60-item measure of receptive vocabulary and range of general information about the world (nature, geography, the arts, science, etc.) (Kaufman & Kaufman, 2004). Each item on this

subtest consists of six pictures. The examinee is given a word or asked a question and the examinee is required to point to the picture that best illustrates the word or answers the question. The Riddles subtest is a 48-item subtest that measures verbal comprehension, reasoning, and vocabulary knowledge. On this subtest, the examiner asks a riddle, and the examinee either points to the picture that shows the answer to the riddle (Items 1-8) or says a single word that answers the riddle (Items 9-48) (Kaufman & Kaufman, 2004).

The Nonverbal score is comprised of one subtest, Matrices, which measures fluid reasoning. The Matrices subtest measures the ability to solve new problems by assessing an individual's ability to perceive relationships and complete visual analogies (Kaufman & Kaufman, 2004). The Matrices subtest is a 46-item nonverbal measure composed of several types of items involving visual stimuli, both meaningful (people and objects) and abstract (designs and symbols) (Kaufman & Kaufman, 2004). On this subtest, examinees have five options to choose from to complete the pattern. The IQ Composite score is calculated using both the Verbal score and the Nonverbal score.

The KBIT-2 was normed on a nationwide sample of 2,120 children and adults aged 4-90 years. The standardization sample was obtained from 113 sites in 34 states and the District of Columbia. The sample was stratified according to the March 2001, Current Population Survey figures on sex, education level, race/ethnicity, and geographic region. The examinees in the standardization sample spoke English, were noninstitutionalized, and did not have physical, perceptual, or psychological impairments that would prevent them from being able to perform the tasks (Kaufman & Kaufman, 2004). The school-aged sample included students in both special education as well as gifted/talented programs.

Reliability for KBIT-2 scores was assessed using internal-consistency reliabilities and test-retest reliability. Internal-consistency reliability was computed for the norm sample using the split-half method. Internal-consistency reliabilities for the Verbal score ranged from .86 to .96 (M = .91). Score reliability for children and adolescents aged 4 through 18 was high (M = .90). For the Nonverbal score, reliabilities ranged from .78 to .93 (M = .88). Once again score reliability was high for children and adolescents 4 through 18 (M = .86). The reliability of the IQ Composite was excellent, ranging from .89 to .96 (M = .93) across the entire age range (Kaufman & Kaufman, 2004).

Test-retest reliability was assessed by administering the KBIT-2 twice to 271 examinees aged 4 through 89. The test-retest interval ranged from 6 to 56 days with an average interval of about 4 weeks. The test-retest reliability for the Verbal score ranged from .88 to .93 (M = .91), and those of the Nonverbal score ranged from .76 to .89 (M = .83). The test-retest reliability for the IQ Composite ranged from .88 to .92 (M = .90). The average increase in the Verbal standard score over the four-week period was 3.6 points, while the increase in the Nonverbal score was 3.3 points. The average increase in the IQ Composite over the four-week period was 4.0 points (Kaufman & Kaufman, 2004).

Several validity studies were conducted between the KBIT-2 and other cognitive test batteries. Validity coefficients between the KBIT-2 and the WISC-III with a sample size of 43 participants showed moderate to high correlations between the two measures. The KBIT-2 Verbal Score correlated highly with the WISC-III Verbal IQ (.83) and the WISC-III Verbal Comprehension Index (.84). However, the KBIT-2 Nonverbal IQ correlated much lower with the WISC-III Performance IQ (.53) and the WISC-III Perceptual Organization Index (.47). The

correlation between the KBIT-2 IQ Composite and the WISC-III Full Scale IQ was .76 (Kaufman & Kaufman, 2004).

Validity coefficients between the KBIT-2 and the WISC-IV with a sample size of 63 participants showed moderate to high correlations between the two measures. The correlation between the KBIT-2 Verbal score and the WISC-IV Verbal Comprehension Index score was .79. However, the correlation between the KBIT-2 Nonverbal score and the Perceptual Reasoning Index score was much lower with a correlation of .56. The IQ Composite of the KBIT-2 correlated highly with the WISC-IV Full Scale IQ with a correlation of .77 (Kaufman & Kaufman, 2004).

The purpose of the current study was to independently evaluate the validity of the recently published KBIT-2 with the recently published WISC-IV. Based on the previous research comparing the K-BIT to the Wechsler scales, the hypotheses were: 1) there would be a moderate to high correlation between the KBIT-2 IQ Composite score and the WISC-IV FSIQ, 2) there would be a moderate to high correlation between the KBIT-2 Verbal IQ score and the WISC-IV VCI, 3) there would be a moderate to high correlation between the KBIT-2 Nonverbal IQ score and the WISC-IV PRI, 4) there would be no significant difference between the mean KBIT-2 IQ Composite score and the mean WISC-IV FSIQ, 5) there would be no significant difference between the mean KBIT-2 Verbal IQ score and the mean WISC-IV VCI, 6) there would be no significant difference between the mean KBIT-2 Nonverbal IQ score and the mean WISC-IV PRI, and 7) there would be no significant difference between the KBIT-2 VIQ-NIQ discrepancy and the WISC-IV VCI-PRI discrepancy.

Method

Participants

The 48 students in the present study were enrolled in kindergarten through ninth grade from school districts in central Illinois and were referred for initial or triennial multidisciplinary evaluation. Seventy-one percent ($n = 34$) of the participants were male and 30% ($n = 14$) were female. The students ranged in age from 6.17 years to 15.67 years, with a mean age of 11.07 years ($SD = 2.93$). Ethnic characteristics of the participants were as follows: Caucasian, 90% ($n = 43$) and Black/African American, 10% ($n = 5$). All students in the present study were English speaking and allowed for the appropriate administration of the tests. Student disabilities were determined through multidisciplinary evaluation teams using federal and state special education criteria. Results of the evaluations indicated that 13 (27%) participants were not disabled, whereas 16 (33%) were learning disabled, 4 (8%) were mildly mentally retarded, 3 (6%) were other health impaired, 2 (4%) were seriously emotionally disabled, and 1 (2%) was speech/language impaired.

Instruments

Kaufman Brief Intelligence Test, Second Edition (KBIT-2).

"The Kaufman Brief Intelligence Test, Second Edition (KBIT-2) 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 years, 0 months through 90 years, 11 months" (Kaufman & Kaufman, 2004, p. 1). The KBIT-2 is composed of three subtests: Verbal Knowledge, Riddles, and Matrices. The KBIT-2 takes approximately 15 to 30 minutes to administer and was standardized on a representative sample ($N = 2120$), closely approximating

the March 2001 Current Population Survey data on variables of gender, geographic region, socioeconomic status, and race/ethnic group.

Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV).

The WISC-IV is an individually administered test of intellectual abilities for children aged 6 years through 16 years, 11 months (Wechsler, 2003). Unlike the previous editions of the WISC that provided three composite scores (VIQ, PIQ, and FSIQ), the latest revision provides a measure of general intellectual functioning (FSIQ) and four index scores (VCI, PRI, WMI, and PSI). The WISC-IV was standardized on a representative sample ($N = 2,200$) closely approximating the 2000 United States Census on age, sex, race, parent education level, and geographic region.

Procedure

In order to obtain participants for this study, school psychologists and school psychology interns in several school districts and special education cooperatives throughout the state of Illinois were contacted. Four school psychologists and three school psychology interns agreed to help in the data collection for this study. School psychologists' and school psychology interns' participation in helping with data collection was voluntary. The school psychologists and school psychology interns were informed of the procedures for collecting the data for this study and were thanked for their participation.

The participants for the study were obtained from referrals for initial psychoeducational evaluations or reevaluations to school psychologists and school psychologist interns. Since the parents of the participants had already provided informed consent for testing, further parental consent was not required for the purposes of this study. All data were anonymously collected and no personally identifiable information was collected. The school psychologists and school

psychology interns recorded only the demographic information and test scores for each participant and no other identifying information was recorded. Therefore, there was no identifying information to link the test scores to the participant.

The school psychologists and school psychology interns were asked to administer the WISC-IV and the KBIT-2 in random counterbalanced order in order to control for order effects. Also, each school psychologist and school psychology intern was asked to complete both test administrations within the same test session. Once both tests had been administered, the school psychologist or school psychology intern completed the test summary sheet and returned the test summary sheets to this researcher.

Results

Descriptive statistics, Pearson product-moment correlation coefficients, and r^2s for the KBIT-2 and WISC-IV are presented in Table 1. Correlations ranged from .29 to .76 with a mean correlation of .55. As expected, the correlation between the KBIT-2 IQ Composite and the WISC-IV Full Scale IQ score was statistically significant, ($r = .76$). Furthermore, correlations between the KBIT-2 VIQ and WISC-IV VCI ($r = .76$) and KBIT-2 NIQ and WISC-IV PRI ($r = .67$) were also statistically significant.

Differences between correlation coefficients were tested using Hotelling's *t*-test formula (Guilford & Fruchter, 1978). As expected the KBIT-2 VIQ had a significantly higher correlation with the WISC-IV VCI than with the WISC-IV PRI, $t(45) = 2.64$, $p < .01$. However, the KBIT-2 NIQ correlated equally well with both the WISC-IV VCI and WISC-IV PRI, $t(45) = 1.48$, *ns*.

Students obtained higher KBIT-2 IQ Composite scores ($M = 91.06$, $SD = 13.77$) than WISC-IV Full Scale IQ scores ($M = 85.27$, $SD = 14.20$), $t(47) = 4.10$, $p < .01$. Table 2 presents descriptive statistics and effect size estimates. The effect size between the KBIT-2 IQ

Table 1

Descriptive Statistics and Pearson Product-Moment Correlations and Shared Variance Between the WISC-IV and KBIT-2

				Kaufman Brief Intelligence Test		
WISC-IV	*n*	*M*	*SD*	Verbal IQ	Nonverbal IQ	IQ Composite
FSIQ	48	85.27	14.20	.64 (.41)	.66 (.44)	.76 (.58)
VCI	48	86.77	13.18	.76 (.58)	.51 (.26)	.74 (.55)
PRI	48	89.77	14.81	.51 (.26)	.67 (.45)	.68 (.46)
WMI	48	87.08	13.83	.29 (.08)	.47 (.22)	.44 (.19)
PSI	48	89.54	16.85	.34 (.12)	.36 (.13)	.40 (.16)
			M	91.44	92.69	91.06
			SD	14.03	13.43	13.77

Note. WISC-IV = Wechsler Intelligence Scale for Children-Fourth Edition, FSIQ = Full Scale IQ, VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index. r^2 presented in parentheses.

Table 2

Descriptive Statistics, 95% Confidence Intervals, Dependent t-tests, and Effect Size Estimates for Composite Comparisons

			95% CI				
	M	*SD*	Lower	Upper	*t*	*p*	Δ
FSIQ	85.27	14.20	8.63	2.95	4.10	.000	.39
IQ Composite	91.06	13.77					
VCI	86.77	13.19	7.43	1.90	3.39	.001	.31
VIQ	91.44	14.03					
PRI	89.77	14.81	6.25	.41	1.76	.085	.19
NIQ	92.69	13.43					
VCI-PRI	3.00	11.72	-5.64	2.14	.91	.370	.12
VIQ-NIQ	1.25	13.98					

Note. FSIQ = Full Scale IQ, VCI = Verbal Comprehension Index, VIQ = Verbal IQ, PRI = Perceptual Reasoning Index, NIQ = Nonverbal IQ.

Composite and the WISC-IV Full Scale IQ was .39, which Cohen (1960) indicated as a medium effect size. Furthermore, students obtained higher KBIT-2 VIQ scores ($M = 91.44$, $SD = 14.03$) than WISC-IV VCI scores ($M = 86.77$, $SD = 13.18$), $t(47) = 3.39$, $p < .01$. The effect size of this comparison was .31, which also indicated a medium effect size. However, students obtained

equivalent KBIT-2 NIQ (M = 92.69, SD = 13.43) and WISC-IV PRI (M = 89.77, SD = 14.81) scores, $t(47) = 1.76$, *ns*.

When comparing the Verbal-Nonverbal discrepancy of the KBIT-2 with the VCI-PRI discrepancy of the WISC-IV, there was no significant difference between the discrepancies, $t(47) = .91$, *ns*. The correlation between the KBIT-2 VIQ-NIQ and the WISC-IV VCI-PRI was .47, $p < .01$. This correlation indicated that only 22% ($r^2 = .22$) of the variability in WISC-IV VCI-PRI discrepancy ($M_{VCI\text{-}PRI}$ = -3.00, SD = 11.72) was accounted for by the KBIT-2 VIQ-NIQ discrepancy ($M_{VIQ\text{-}NIQ}$ = -1.25, SD = 13.98).

Discussion

The present study compared KBIT-2 and WISC-IV scores in a sample of kindergarten through ninth grade students who were referred for a multidisciplinary evaluation. The high statistically significant correlation ($r = .76$) between the KBIT-2 IQ Composite and the WISC-IV Full Scale IQ scores supported the construct validity of the KBIT-2 and its value as a brief screening measure of general intelligence. Furthermore, the correlation obtained between the KBIT-2 IQ Composite and the WISC-IV Full Scale IQ in the present study is consistent with the correlation obtained in the subsample of the KBIT-2 standardization sample ($r = .77$, Kaufman & Kaufman, 2004).

When examining the comparison between the KBIT-2 VIQ and the WISC-IV VCI, the correlations obtained were also very consistent. In the present study the correlation between the two scores, $r = .76$, was nearly identical to the correlation obtained in the subsample of the KBIT-2 standardization sample ($r = .79$, Kaufman & Kaufman, 2004).

When examining the comparison between the KBIT-2 NIQ and the WISC-IV PRI, the correlations obtained were again consistent. In the present study the correlation between these

two scores was $r = .67$ compared to $r = .56$ for the KBIT-2 standardization sample (Kaufman & Kaufman, 2004). No statistical difference was found between these correlations, $Z = .90$, *ns*. The moderate to high correlations in the present study were consistent with past studies on the K-BIT and the WISC-III (Canivez, 1995; Canivez et al, 2005; & Chin et al, 2001).

In the current study, the KBIT-2 tended to yield higher scores in the areas of verbal, nonverbal, and overall cognitive ability compared to the WISC-IV. The KBIT-2 VIQ score was, average, 4.67 points higher than the WISC-IV VCI. However, Kaufman and Kaufman (2004) reported that the KBIT-2 VIQ was only 1.8 points higher than the WISC-IV VCI. In the current study, the KBIT-2 NIQ score was, average, 2.92 points higher than the WISC-IV PRI, whereas, Kaufman and Kaufman (2004) reported the WISC-IV PRI was 1.6 points higher on average than the KBIT-2 NIQ. Finally, in the current study, the KBIT-2 IQ Composite score was on average 5.79 points higher than the WISC-IV FSIQ, whereas, Kaufman and Kaufman (2004) reported the IQ Composite score was on average 1.3 points higher than the WISC-IV FSIQ. Therefore, the present study indicated that the KBIT-2 VIQ and IQ Composite tended to yield higher scores on cognitive ability compared to the results obtained by Kaufman and Kaufman (2004) in a subset of the KBIT-2 standardization sample. The higher KBIT-2 IQ Composite compared to the lower WISC-IV FSIQ is consistent with past research on the K-BIT and the WISC-III (Canivez, 1995; Canivez, 2005; Chin et al, 2001; Donders, 1995; Grados & Russo-Garcia, 1999; Prewett, 1995; Thompson et al, 1997). However, the effect sizes in the current study were higher than in previous studies.

When examining the KBIT-2's ability to assess a verbal-nonverbal discrepancy as is done with the WISC-IV VCI-PRI discrepancy, no statistically significant difference was shown in the average discrepancy, suggesting that the KBIT-2 VIQ-NIQ discrepancy and WISC-IV VCI-PRI

discrepancy were similar. However, only 22% of the variability in WISC-IV VCI-PRI discrepancy was accounted for by the KBIT-2 VIQ-NIQ discrepancy. This is consistent with the low to moderate magnitude obtained by Canivez (1995) and Canivez (2005). These results further emphasize that the KBIT-2 VIQ-NIQ discrepancy appears to be an inadequate predictor of the WISC-IV VCI-PRI discrepancy.

There were several limitations to the present study. First, the sample size in the current study was small. The small sample size produced somewhat less variability in the scores. Second, the students in the current study were referred for psychoeducational evaluations. Several of the students in the study were receiving reevaluations and had already been identified with some type of disability. Third, the participants in the study consisted of only Caucasian and African American students. There were no Hispanic, Asian, Indian, or other ethnicities included in this study. Fourth, all students included in the study were from central Illinois. All of these factors limit the generalizability of the results of this study.

Future research should be conducted on the KBIT-2 and WISC-IV as well as other comprehensive intellectual ability measures. This research should focus on normal individuals as well as students with different ethnicities, social economic status, and disabilities to further define the KBIT-2's psychometric characteristics. Future research should also focus on obtaining a larger sample size and having participants from different geographic locations. Research should also be conducted using the WISC-IV, KBIT-2, and a standardized achievement test on students with a learning disability who are receiving triennial reevaluations. This research would be conducted to evaluate the diagnostic efficiency of the KBIT-2 when the KBIT-2 is being used as a cognitive measure to assess an ability-achievement discrepancy in reevaluation situations as in Canivez (1996).

Results from this study suggested that caution is needed in utilization of the KBIT-2 in the assessment of children for psychoeducational evaluations. The KBIT-2 may be useful in reevaluations where a brief measure of cognitive ability can be utilized. Also the KBIT-2 VIQ-NIQ discrepancy was not an accurate predictor of the WISC-IV VCI-PRI discrepancy, which was consistent with the results obtained by Canivez (1995, 2005). Although the KBIT-2 may be useful for reevaluation purposes, research, and screening purposes, it is not recommended that the KBIT-2 be used in initial psychoeducational evaluations. A comprehensive ability measure most likely needs to be used in these cases.

References

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.

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). 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., Neitzel, R., & Martin, B. E. (2005). Construct validity of the Kaufman Brief Intelligence Test, Wechsler Intelligence Scale for Children-Third Edition, and Adjustment Scales for Children and Adolescents. *Journal of Psychoeducational Assessment, 323,* 15-34.

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.

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

Donders, J. (1995). Validity of the Kaufman Brief Intelligence Test (K-BIT) in children with traumatic brain injury. *Assessment, 2,* 219-224.

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.

Grados, J. J., & Russo-Barcia, K. A. (1999). Comarison 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.

Guilford, J. P., & Fruchter, B. (1978). Fundamental statistics in psychology and education. New York: Mc Graw-Hill.

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.

Kaufman, A. S., & Kaufman, N. L. (1990). Manual for the Kaufman Brief Intelligence Test. 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.

Kaufman A. S. & N. L. Kaufman (2004). Kaufman brief intelligence test-second edition. Circle Pines, MN: AGS Publishing.

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*, 977-982

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., & McCaffery, L. K. (1993). A comparison of the Kaufman Brief Intelligence

Test (K-BIT), the Stanford-Binet Intelligence Scale (4th ed.), a 2-subtest short form, and the Kaufman Test of Educational Achievement (K-TEA). *Psychology in the Schools, 30,* 299-304.

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

Slate, J. R. & Graham, L. S. (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 Psychoeducational Assessment, 12,* 331-350.

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.

Webber, L. S. & McGillivray, J. A. (1998). An Australian validation of the Kaufman Brief Intelligence Test (K-BIT) with adolescents with an intellectual disability. *Australian Psychologist, 33,* 237-237.

Wechsler, D. (1939). Manual for the Wechsler-Bellevue Intelligence Scale. New York: The Psychological Corporation.

Wechsler, D. (1949). Manual for the Wechsler Intelligence Scale for Children. New York: The Psychological Corporation.

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

Wechsler, D. (1976). Manual for the 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. (1997). Manual for the Wechsler Adult Intelligence Scale, Third Edition. San Antonio, TX: The Psychological Corportation.

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

Wechsler, D. (2002). Manual for the Wechsler Preschool and Primary Scale of Intelligence, 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.