Validity of the K-BIT and a WISC-III Short Form with a Sample of Adolescent Young Offenders

JANET L. BROWNE ©

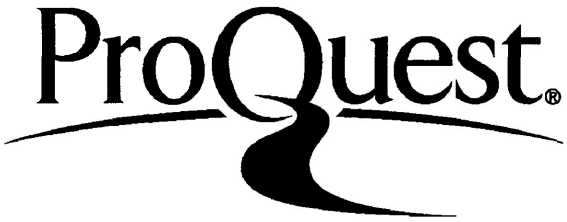
Lakehead University

Thesis Supervisor: Anthony P. Thompson, Ph.D.,C.Psych

A Thesis Submitted to the Department of Psychology

in Partial Fulfilment of the Requirements for the Masters of Arts Degree

in Clinical Psychology



ISBN 0-612-09196-1

Acknowledgements

To begin, I would like to thank Dr. Tony Thompson for all of his encouragement, time, and guidance. Your expertise and continual support have been inspiring. Thank you for being a fabulous advisor; I am grateful to have had the opportunity to work with you. I would also like to thank Dr. Fred Schmidt who worked as a second reader and clinical advisor on this project. Your ongoing enthusiasm, flexibility, and involvement meant a great deal to me - I appreciate that you always had time to listen to my academic, research and career concerns. Thank you for all of your help and all of your guidance! I would also like to thank Marian Boer for her enormous contribution to this project; I appreciate all the time and effort that you afforded. I would also like to extend my gratitude to Dr. Dwight Mazmanian for sitting on my committee as the internal examiner; thank you very much for the meticulous review and thoughtful critique. I would also like to thank Dr. Steve LoBello for lending his expertise as the external examiner for this project.

Thank you to LRFC and William W. Creighton Youth Services for making this project possible. A special thanks to all of the staff at William W. Creighton Youth Services for being so helpful and friendly, and for making me feel very welcome.

Thank you to Brian Miller, for standing by me and for making our relationship work despite the distance. I appreciate your support and love. Thank you also to my friends - for all the laughter, rides, support and critique. Finally, and perhaps most of all, I would like to thank all of the young people who participated in this study - you have my best wishes for the future.

Table of Contents

List of Tables

Abstract

The present study investigated the concurrent validity and test-retest reliability of the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990) by comparing it to the full scale administration of the Wechsler Intelligence Scale for Children third edition (WISC-III; Wechsler, 1991) in a clinical trial with 35 young offenders. Using a test-retest methodology, the concurrent validity of a four subtest short form of the WISC-III (SF4) was also assessed.

Results of this study show that while the K-BIT is a reliable measure, it did not perform well in terms of validity criteria typically applied to short form estimates of intelligence. While mean Full Scale IQ (FSIQ) and Performance IQ (PIQ) were not significantly different, the K-BIT overestimated Verbal IQ (VIQ) by an average of five points. Correlation coefficients between the K-BIT and WISC-III were significant but not impressive and only 40% to 50% of the IQ estimates were within WISC-III 95% confidence limits. On the other hand, SF4 performed reasonably well as an estimate of FSIQ, with a high correlation coefficient and a low misclassification rate. When administered first SF4 overestimated FSIQ by three IQ points, however when administered following the full scale, SF4 overestimated FSIQ by six IQ points.

Validity of the K-BIT and a WISC-III Short Form with a Sample of Adolescent Young Offenders

Intelligence testing is frequently an integral part of the psychological assessment process. Intelligence test results provide a context for understanding aspects of cognition, personality and social functioning, and may indicate a need for further testing (Haynes, 1983). Classic intelligence measures typically involve lengthy assessment and scoring procedures. Short forms of traditional measures along with brief new intelligence tests may provide relatively quick assessments of global intelligence that are still valid and reliable. Such scales would have clinical value as a general measure of intelligence in cases where the intellectual assessment is not the focal issue, or as a screening tool to indicate when further assessment is warranted. The current study was designed to assess the validity and reliability of a new, brief psychometric test of intelligence, the Kaufman Brief Intelligence Test (Kaufman & Kaufman, 1990) and the validity of a four-subtest short form of a standard intellectual instrument, the Wechsler Intelligence Test for

Children - Third Edition (Wechsler, 1991), with a sample of young offenders.

Wechsler Intelligence Scales

While there are many intelligence tests available, the Wechsler scales (Wechsler, 1981; 1989; 1991) currently dominate the field (Naugle, Chelune, & Tucker, 1993; Watkins, 1986). Collectively, these scales have shaped the clinical definition of intelligence (Leckliter, Matarazzo & Silverstein, 1986; Matarazzo, 1972), while inspiring the development of new psychometric measures (Kaufman & Kaufman, 1990). Current versions of the Wechsler scales are the Wechsler Adult Intelligence Scale - Revised (WAIS-R; Wechsler, 1981), the Wechsler Intelligence Scale for Children - Third Edition (WISC-III; Wechsler, 1991) and the Wechsler Preschool and Primary Scale of Intelligence - Revised (WPPSI-R; Wechsler, 1989).

The Wechsler scales were designed to partition intelligence into two qualitatively distinct domains. Consequently, each Wechsler scale is comprised of a variety of subtests which predominantly measure either verbal or performance (non-verbal) intelligence. Hence each scale produces a Verbal Scale score (VIQ), a Performance Scale score (PIQ)

and an overall composite or Full Scale score (FSIQ). The dichotomy of performance versus verbal intelligence is supported by factor analytic research and has strong diagnostic utility (Matarazzo, 1972; Wechsler, 1974;1981). These tests have been shown to yield valid and reliable measures of intelligence for their targeted age groups (Sattler,1992; Wechsler, 1974; 1981; 1991).

Despite their outstanding psychometric properties, the Wechsler scales are time consuming to administer as each version consists of ten or eleven standard subtests and several optional supplemental subtests. Sixty to ninety minutes are typically required for complete administration (Ryan, 1983; Wechsler, 1981); a process which is becoming increasingly expensive. In situations where a global estimate of intelligence is sufficient or during screening evaluations, psychologists have begun to administer abbreviated forms of the WISC-III, WAIS-R, and the WPPSI-R (Fell & Schmidt-Fell, 1982; Ryan, 1981; Watkins, 1983). These short forms are intended as a screening device and in selected cases may be followed by full scale administration or other psychometric scales (Kaufman, 1990). Thompson and LoBello (1994) have provided some psychometric criteria to help identify when further intellectual assessment may be warranted.

Types of Short Forms

Short forms of the various Wechsler intelligence scales typically involve a reduced number of subtests or a reduced number of items within each subtest. The latter procedure is termed the "selected item approach" (Satz & Mogel, 1962) and carries the advantage of representing all subtests from the full Wechsler scale. Satz-Mogel short forms also tend to preserve the high validity coefficients of the full scale test more so than selected subtest versions (Adams, Smigielski, & Jenkins, 1984; Watkins, 1986). However, research has shown that administration time is not significantly reduced by the Satz-Mogel short forms whereas selected subtest short forms do take much less time to administer (Watkins, 1986). Silverstein (1990) has concluded that the best combinations of four or five subtests consistently yield coefficients comparable to those achieved from the selected item approach. Furthermore, short form reliability was superior in particular selected subtest abbreviations (Ryan, 1981; Silverstein, 1982; 1990; 1991). The reduced subtest approach also allows for follow-up testing via administration of the remaining scales. Satz-Mogel short forms preclude this option.

Evaluating the Validity of Subtest Short Forms

Clinical trials are frequently reported to assist practitioners in the use and selection of subtest short forms. These clinical trials typically report on the validity of the short form version as it is applied to particular clinical samples (see selected reviews by Kaufman, 1990; Thompson, Howard, & Anderson, 1986; Watkins, 1986.) Such information is useful and necessary for clinicians as subtests are selected for abbreviated testing based on their psychometric properties in the standardization sample. Hence clinicians need to know whether these promising combinations generalize to clinical samples.

The validational criteria used to assess short form intelligence measures have usually been those proposed by Resnick and Entin (1971). Specifically, (a) the difference between the mean IQ of the short form (SFIQ) and the mean full scale measure (FSIQ) must be small and statistically nonsignificant, (b) the correlation between SFIQ and FSIQ must be significant and account for a substantial percentage of the variance shared by the two measures, and finally (c) there must be a high correspondence between the SFIQ classification and the classification based on the FSIQ.

Results from clinical studies with adults, adolescents, and children have been mixed. Some studies have supported the clinical utility of short forms (e.g., Donders, 1992; Dumont & Faro, 1993; Fell & Schmidt-Fell, 1982), while other studies question their validity and suggest caution (e.g., Cyr & Atkinson, 1991; Ryan 1981; Watkins, 1986; Watkins, McKay, Parra, & Polk, 1987). In terms of the Resnick and Entin criteria, results generally meet the first criterion, show mixed results on the second and fall short on the third (Silverstein, 1985a; 1985b; Thompson, 1987; Watkins, 1986). However, there has been some debate concerning the value of the Resnick and Entin criteria (see summaries by Silverstein, 1982; Thompson, 1987; Watkins, 1986). For example, classification mismatches can occur when SFIQ and FSIQ differ by only one point, (e.g., 89 vs 90) but large differences (e.g., 90 vs. 109) do not always result in misclassification (Silverstein, 1985b; Thompson, 1987). Thompson (1987) suggested that a better criterion would be for SFIQ to fall within FSIQ confidence limits for a large proportion of the short form estimates. Silverstein (1985b) argues that all three of these validational criterion lack legitimacy. In his review of short form research, Silverstein (1985b) found that there will always be a high correlation between the FSIQ and the SFIQ if the sample is large and representative since the

short forms were initially chosen based on this criteria. Further, Silverstein (1985b) shows that even trivial discrepancies between the SFIQ and FSIQ will achieve statistical significance if the sample size is large enough. The third criterion is also considered unreasonable, because the aim of the IQ measurement is not merely to classify. Therefore Silverstein (1989) argues that correlation coefficients from the standardization data are the only salient criterion to evaluate short form validity. However, it seems prudent for clinicians to have as much information as possible about short form performance, and mixed clinical trials with short forms suggest that further investigation is warranted. Consequently the tradition of evaluating short forms according to the Resnick and Entin criteria continues.

Methodological Issues with Subtest Short Forms

The evaluation of short forms has evolved from reexamination of the standardization data to application in clinical trials. Although several studies of the earlier versions of the Wechsler children's scale have appeared (e.g., Fell & Schmidt-Fell, 1993; Prewett, 1992a; 1992b; Resnick & Entin, 1971; Ryan, 1981; Silverstein, 1975; Watkins, 1986), the author is aware of only one study that involves the WISC-III (Dumont,

R., & Faro, C., 1993). Also until recently, few clinical validity trials had actually administered short forms as separate tests. Almost exclusively investigators have relied on administration of the full scale. The relevant short form subtests have then been extracted and used to estimate IQ which has been compared with Full Scale IQ. This trend is changing though, with attention being paid to the conditions unique to bona fide short form administration. As Thompson and his colleagues noted (1986), abbreviated intellectual testing compared to full scale examination involves reduced levels of effort, attention and motivation on the part of the examiner and examinee alike. Consequently, evaluation of subtest short forms which have been administered as part of the full scale may be producing inaccurate estimates of the short form validity. Thompson (1987) has proposed two alternate methodologies for short form administration in clinical studies; either the short form can be administered first, followed by the remaining subtests in order, or subjects can be given both the short form and full scale at different times.

These methodologies are beginning to be reported in the literature (Plumridge, unpublished thesis, 1994; Thompson, 1995; Thompson, Howard & Anderson, 1986). Thompson and his colleagues (1986)

examined two- and four-subtest short forms of the WAIS-R using a sample of ninety (45 male and 45 female) psychiatric inpatients. Their study involved three administrative conditions: (a) 15 males and 15 females received the SF2 first, then the remaining WAIS-R subtests in standard order, (b) 15 males and 15 females received the SF4 first, followed by the remaining subtests in standard order, and (c) 15 males and 15 females received the full WAIS-R in standard order. This methodology allowed validity to be compared for short forms administrated under conditions which approximate intended clinical use versus the expedient method of re-scoring a full administration. The study revealed that the SF2 yielded a greater overestimate of IQ when it was administered first. The SF4 was not as susceptible to this methodological effect. These results demonstrate the importance of looking at different administrative conditions when evaluating the validity of Wechsler short forms.

Plumridge (1994, unpublished thesis) investigated several subtest short forms of the WAIS-R using a sample of normal subjects. Similar to the above study (Thompson et al., 1986), she examined the accuracy of short form IQ estimates by comparing short forms extracted from the full scale to short forms that were administered alone followed by the

remaining WAIS-R subtests. For the four subtest short form, Plumridge found that SFIQ and FSIQ were not significantly different, regardless of the way SF4 was administered. For SF2, the results showed that SFIQ and FSIQ were not significantly different when the short form was administered first followed by the remaining subtests. However, when the two subtest short form was rescored from the standard full scale administration the SFIQ was significantly larger than FSIQ. Consequently like Thompson et al. (1986), Plumridge found that the SF2 was susceptible to the condition of administration, although the direction of influence was in the opposite direction (ie. SF2 administered first versus rescored from full administration).

Thompson (1995) investigated the validity of a WAIS-R short form using a test-retest methodology, again administering the short forms under conditions that were similar to their intended clinical use. Administration was between two sessions such that 21 participants received the SF4 first, and 26 received the full WAIS-R first. Average test-retest period was 54 days (range 14 to 126 days). When administration of the SF4 followed full scale administration of the WAIS-R, the short form yielded an average overestimate of 11 points. However, average IQ overestimate of SF4 when administered first was

only 2.38 IQ points which was not a significant difference. Consequently when the SF4 is administered first it is a reasonable, but not interchangeable, estimate of FSIQ.

It is important to note that the complexity of short form validity would not have been revealed with traditional rescoring methodology. Consequently, validity studies using methodologies alternate to rescoring are helping to elucidate further the clinical utility of short forms. Clearly, such studies with abbreviated versions of the WISC-III would be welcome.

New Forms of Brief Intelligence Testing

In addition to abbreviated versions of the Wechsler scales, Kaufman and Kaufman have developed a new, brief intelligence scale called the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990). The K-BIT yields an estimate of full-scale IQ much like Wechsler short forms. In addition, the K-BIT provides an estimate of verbal and performance IQ. The K-BIT is motor free; is applicable for ages 4 to 90 years, and takes only 15 to 30 minutes to administer in full. Three subtests are included in the K-BIT; Expressive Vocabulary, Definitions and Matrices. The first two subtests are used to estimate the Verbal IQ.

The third subtest requires visuospacial reasoning through nonverbal analysis to provide an estimate of Performance IQ. The K-BIT was standardized on a sample of 2,022 persons with a racial-ethnic representation proportional to a national census of the United States. The standardization sample also accurately represented the geographical regions and the socioeconomic strata of the United States.

Initial data on the psychometric properties of the K-BIT seem promising. The K-BIT manual reports an internal consistency coefficient of .92, with split-half reliability coefficients ranging (for different age groups) from .89 to .98 for Vocabulary (expressive vocabulary and definitions), .74 to .95 for the Matrices subtest and from .88 to .98 for the IQ Composite. The test-retest reliability coefficients were calculated for a sample of 232 normal children and adults (range 5 to 89 years) with a mean retest interval of 21 days (range 12 to 145 days). The reported overall mean retest reliability coefficient (averaged across age groups) was .94 for Vocabulary (range .86 to .97), .85 for Matrices (range .80 to .92), and .90 for the K-BIT IQ Composite (range .92 to .95).

The K-BIT's standard errors of measurement (SEM) are also reported for various age groups. On average, K-BIT IQ Composite and Vocabulary standard scores have SEMs of approximately 4 points, while

the Matrices' standard scores have SEMs of about 5.5 points. Moderate intercorrelations between the Verbal and Performance standard scores are reported, showing an increasing relationship with age. For ages 8 to 19 years, the average value was $\underline{r}$ = .60. The authors of the test state that these intercorrelations are low enough to suggest unique contributions of each subtest to the K-BIT IQ Composite score.

Construct and concurrent validational studies were conducted during the standardization of the K-BIT. Concurrent validity data were based on 20 studies which included 982 children, adolescents and adults. The K-BIT IQ scores were compared to other intellectual measures, including the Verbal, Performance and Full Scale IQ scores of the WISC-R (Kaufman & Kaufman, 1990). Of particular relevance to the current study are results for the adolescent sample. Thirty-five normal children and adolescents between the ages of 6 and 15 years (mean age = 10 years, 8 months) were given both the K-BIT and WISC-R, and the correlation coefficients were examined. No further details of the method were provided in the K-BIT manual. The relationship between the K-BIT Vocabulary subtest standard score and WISC-R Verbal IQ yielded a correlation coefficient of .78. The means for the K-BIT and WISC-R on the Vocabulary scale were 106.0 and 110.2 respectively (4.2 mean score

difference). The correlation coefficient reported for the K-BIT Matrices subtest standard score and the WISC-R Performance IQ score was .50. The means for the K-BIT Matrices and WISC-R Performance were 104.1 and 109.9 respectively (5.8 mean score difference). Finally, the WISC-R Full Scale IQ and K-BIT Composite IQ scores achieved a correlation coefficient of .80. The mean K-BIT Composite IQ was 105.5, and the mean FSIQ was 111.4 (5.9 mean score difference). The manual did not report whether these differences were significant. It is important to note that because the WISC-III was in the process of standardization at the time of the K-BIT publication, concurrent validation data with the new version are not available in the K-BIT manual.

Other investigators have begun to evaluate the concurrent validity of the K-BIT. Prewett (1992a) investigated the relationship between the K-BIT and the WISC-R with forty academically deficient male juvenile delinquents ranging from 12 to 16 years of age. The WISC-R and the K-BIT were administered in counterbalanced order in a single session. Overall, the mean K-BIT Composite IQ (82.3) did not differ significantly from the mean FSIQ scores (82.7) obtained on the WISC-R ($t(39) = .34$, $p > .74$). The mean Verbal IQ estimate of the K-BIT (81.4) was 4.0 points higher than the mean WISC-R Verbal IQ (77.4). The Performance IQ

estimate of the K-BIT (86.7) was 4.7 points lower than the mean WISC-R Performance IQ (91.4). Prewett did not indicate if differences on these scales were significant. Correlation coefficients were r = .70 between Verbal IQs, r = .29 between Performance IQs and r = .64 for Overall IQs. Only the correlation between the Performance IQs was not significant.

Prewett (1992b) went on to investigate the validity of the K-BIT, among a sample of 35 students with a mean age of 10 years, 8 months (age range of 7 to 16 years) who were referred for psycho-educational evaluation. As a group, these students achieved intellectual levels on the WISC-R that averaged one full standard deviation below the mean. Prewett found a substantial correlation between the K-BIT and WISC-R scores when the measures were administered in a single session with test order counterbalanced. Specifically, correlational coefficients obtained were r = .83 for VIQ, r = .70 for PIQ and r = .81 for FSIQ. This study also revealed that the K-BIT IQ Composite significantly underestimated the WISC-R FSIQ (with Overall IQ scores yielding a 6.2 mean score difference). Prewett (1992b) attributed the under-estimate to the 8.3 point discrepancy between the K-BIT Matrices and the WISC-R Performance IQ scores (mean IQ scores obtained were 78.9 and 87.2 respectively). The Verbal IQ scores were not significantly different.

Prewett concluded that the K-BIT may have the tendency to over-identify students who warrant a full evaluation for diagnosis of mental retardation. The author concluded that he supported the use of the K-BIT as a screening instrument, recommending that the WISC-R be used as the comprehensive, follow-up measure of intellectual abilities.

Naugle, Chelune and Tucker (1993) administered the K-BIT and the WAIS-R in the same session to a sample of 200 clinical patients ranging from 16 to 74 years. They did not report if administration order was counterbalanced. Correlation coefficients were $\underline{r}$ = .83 between Verbal IQs, $\underline{r}$ = .77 between Performance IQs and $\underline{r}$ = .88 for Overall IQs. The scores obtained from the K-BIT were comparable to the WAIS-R, but significantly higher. Specifically, the mean Verbal IQ estimate of the K-BIT (98.3) was 3.11 points higher than the mean WAIS-R Verbal IQ (95.2). The Performance IQ estimate of the K-BIT (98.5) was 5.2 points higher than the mean WAIS-R Performance IQ (93.3). Similarly the K-BIT Composite IQ (98.2) was 4.3 points higher than the WAIS-R FSIQ (93.9).

In his most recent study, Prewett (1995) examined the K-BIT and the WISC-III in a sample of fifty students with a mean age of 9 years and 3 months, (range 6 to 14 years of age) referred for psychoeducational

evaluation due to inadequate academic progress. In this single-session, counterbalanced design, Prewett found that the K-BIT tended to yield scores that were significantly higher than those obtained on the WISC-III. Means for K-BIT composite, and WISC-III FSIQ were 76.9 and 72.1 respectively. Prewett did not report comparisons between the Verbal and Performance scales.

Validity is a cumulative enterprise. Results of the K-BIT validity studies are summarized in Table 1. Although the K-BIT scores are comparable, there have been significant over- and under-estimates with different samples and different Wechsler scales. It is not generally reported whether or not these differences have resulted in misclassification of the study's participants. Overall, the results of the reported validational studies have shown consistently high correlations between the K-BIT and Wechsler scales. Clearly further studies are needed to determine the K-BIT's performance with different clinical groups. Also, it would be useful to know how the K-BIT compares with subtest short forms from the newest Wechsler tests.

Table 1. Prior Research Comparing the K-BIT with various Wechsler Intelligence Scales.

Study	Sample[1]	Design	Results (K-BIT)
K-BIT Manual (Kaufman & Kaufman, 1990)	35 N (6-15 yrs)	K-BIT vs. WISC-R no info. on design	< FSIQ (5.9 pts) < PIQ (5.8 pts) < VIQ (4.2 pts)
Prewett (1992a)	40 YO (12-16 yrs)	K-BIT vs. WISC-R counterbalanced single session	FSIQ no difference > VIQ (4.0 pts)[2] < PIQ (4.7 pts)[2]
Prewett (1992b)	35 referred students (PE) (7-16 yrs)	K-BIT vs. WISC-R counterbalanced single session	< FSIQ (6.2 pts) < PIQ (8.3 pts) VIQ no difference
Naugle et al (1993)	200 referred patients (NPE) (16 - 74 yrs)	K-BIT vs. WAIS-R single session - not known if counterbalanced	> FSIQ (4.27 pts) > PIQ (5.19 pts) > VIQ (3.11 pts)
Prewett (1995)	50 referred students (PE) (6-14 yrs)	K-BIT vs. WISC-III counterbalanced single session	> FSIQ (4.8 pts) - no information VIQ or PIQ

[1]N = normal, YO = young offender, PE = psychoeducational examination, NPE = neuropsychological examination

[2] author did not indicate whether the difference was significant or not

Purpose of this Study

The purpose of the current investigation was to assess the concurrent validity of the K-BIT by comparing it to the full scale administration of the WISC-III in a clinical trial with young offenders. The concurrent validity of a four subtest short form of the WISC-III was also assessed with the same sample using the test-retest methodology. Finally, the current investigation also assessed the test-retest reliability of the K-BIT.

Method

Participants

Thirty five young offenders (29 males and 6 females) serving dispositions at a custodial youth services centre participated in this study. The mean age of subjects at time of first testing was 15.12 years (SD = 1.7, range 13 to 16 years). The sample was predominantly Caucasian subjects (74%) with a mean age of 15.08 years and Native Canadian subjects (23%) with a mean age of 15.34 years. Only those participants who had not received intelligence testing in the last two years were recruited.

Measures

The Kaufman Brief Intelligence Test (K-BIT). The K-BIT (Kaufman & Kaufman, 1990) is comprised of three subtests, and yields a Verbal IQ, Performance IQ and Composite IQ score. Verbal IQ is derived from two subtests; Expressive Vocabulary (45 items) which requires naming of pictured objects and is administered to all subjects; and Definitions (37 items) which requires the individual to provide the word that best fits the description and partial spelling of the word, and is administered to subjects 8 years and older. The two Vocabulary raw scores are summed and converted to a standard score (M=100, SD=15) via the tables provided. The third subtest, Matrices, has 45 multiple-choice matrix analogy items which yield a non-verbal or performance score (M=100, SD=15). The sum of the Vocabulary and Performance standard scores is converted to a total test score called the IQ Composite (M=100, SD=15). The K-BIT is motor free; is applicable for ages 4 to 90 years, and takes only 15 to 30 minutes to administer in full.

The Wechsler Intelligence Scale for Children, Third Edition (WISC-III). The WISC-III (Wechsler, 1991) is a well established measure of intellectual potential for persons under the age of 17 years. The test

consists of ten subtests and yields a Verbal, Performance, and Full Scale IQ score. The three available supplementary subtests were not administered in this study.

The Wechsler Intelligence Scale for Children, Third Edition, Four Subtest Short Form (WISC-III SF4). Four subtests of the WISC-III were administered together as a short form. The subtests were the Picture Completion, Information, Block Design and Vocabulary subtests (PC-I-BD-V) and were chosen from Sattler (1993). This particular combination was chosen because previous research has shown that four subtest short forms seem to offer the best estimates of full scale IQ, and because two verbal and two performance subtests have been recommended for subtest short forms (Kaufman, 1991).

Procedure

Ethical approval for this project was obtained from the Youth Services Centre, the Lakehead Regional Family Centre, and the Lakehead University Ethics Advisory Committee to the Senate Research Committee (see Appendix A). Letters of introduction which explained the study, participation and feedback were provided to each youth and their parent/legal guardian (see Appendix B). Informed written consent was

obtained from each youth and parent (see Appendix C).

There were two testing sessions for each subject. A counter-balanced and randomized design was attempted. For test-retest reliability, the K-BIT was administered in each session. To evaluate short form validity, the WISC-III short form was administered in one session and the full WISC-III in the other session. Thus, one session required K-BIT and WISC-III SF4, the other K-BIT and WISC-III full. The testing order was counterbalanced within and across testing sessions such that there were eight combinations of test order in total.

Each person who became available for participation was assigned to a particular test order drawn at random from the test combinations in blocks of eight. However, only about two thirds of the participants were randomly assigned as intended. Some subjects left the facility prior to their second testing session due to transfer or early release. Also, time constraints for the project required targeting specific test order combinations to optimize the distribution of participants. In doing so, six subjects completed only the full WISC-III in the first session and the K-BIT and SF4 in the second testing session.

The majority of subjects were tested by the author (66%), the remainder (34%) were tested by two other administrators (one M.A. level

and one Ph.D.). All subjects were tested and retested by the same person. The mean interval between testing sessions was 17.00 days (range 4 to 36 days).

The author checked all raw scores, scale scores and IQ computations for each test, and made any necessary corrections. Subjects received verbal feedback about their performance on the full administration of the WISC-III. A feedback protocol for feedback was used as much as possible to ensure consistent and accurate feedback (see Appendix D).

Results

Descriptive Data.

Based on the full WISC-III administration (n = 33), the mean Verbal IQ (VIQ) was 89.18 (SD = 10.95, range 60 - 110); the mean Performance IQ (PIQ) was 99.79 (SD = 15.56, range 58 - 129); and the mean Full Scale IQ (FSIQ) was 93.45 (SD = 11.84, range 65 - 113). For the entire sample, based upon the first administration of the K-BIT (n = 35), the mean Verbal IQ was 94.57 (SD = 11.02, range 68 - 120); the mean Matrices IQ was 95.46 (SD = 15.64, range 57 - 115); and the mean Composite IQ score was 94.31 (SD = 12.56, range 66 - 118).

K-BIT vs. WISC-III.

Thirty-three participants completed both the K-BIT (first administration) and full WISC-III. The data from one participant were not included because the K-BIT minus WISC-III estimate for this subject (63 IQ points) was more than three standard scores away from the mean in the distribution of difference scores. Consequently this subject was considered an outlier and as such was likely to unduly influence the small data set. Of the remaining thirty-two participants, seventeen had the K-BIT first followed by the WISC-III, and fifteen had the WISC-III followed by the K-BIT. For 17 of these participants, the tests were administered in one session. For the remaining 15 subjects, two sessions were used and the mean retest interval was 16.44 days (range 4 to 36). The distribution of test order by number of testing sessions was not significant at traditional levels (χ^2 , 1df = 2.08, p = .15). Table 2 shows summary IQ statistics for the K-BIT and WISC-III by administration order. Correlation coefficients obtained were r = .73, p < .01 for Verbal IQ; r = .53, p < .05 for Performance IQ; and r = .69, p < .01 for Full Scale IQ. The scores were analyzed as a repeated measures ANOVA comparing IQ results for the two tests and taking into account the order of administration. The administration order was the between-subjects

Table 2. Comparisons of Full Scale WISC-III and K-BIT by Administrative Order.

Administration order	K-BIT (n = 17) *M*	*SD*	WISC-III (n = 15) *M*	*SD*
K-BIT - WISC-III (n = 17)				
Verbal	97.65	9.35	92.88	9.29
Performance[a]	97.59	13.69	98.82	13.04
Full[b]	97.12	11.39	95.00	10.31
WISC-III - K-BIT (n = 15)				
Verbal	90.73	12.54	85.00	11.84
Performance[a]	94.80	15.62	99.53	18.13
Full[b]	91.87	12.82	91.00	13.47

[a] Performance = K-BIT Matrices and WISC-III Performance scales
[b] Full = K-BIT Composite and WISC-III Full Scale

factor, and the test (WISC-III vs. K-BIT) was the within-subjects factor. For Verbal IQ indices, the results showed a significant test effect ($F(1,30) = 12.53$, $p < .001$) in the direction of a larger K-BIT Verbal IQ (means = 89.19 versus 94.41), and a significant between subjects/order effect ($F(1,30) = 4.44$, $p < .044$). The significant between subjects/order effect was in favour of the participants who were administered the K-BIT then WISC-III versus WISC-III then K-BIT. The collapsed marginal IQs were approximately seven points higher for Verbal IQ. For Performance IQ indices, the results showed that there were no significant differences. Comparison of FSIQ versus K-BIT IQ Composite also showed no significant differences.

For K-BIT Verbal scores, 50% were within 95% confidence limits of WISC-III VIQ; 44% of K-BIT Performance scores were within 95% confidences limits of WISC-III PIQ and 46% of K-BIT Composite IQ scores were within the 95% confidence limits of WISC-III FSIQ. Using the Wechsler (1991) intelligence categories, 56% of participants were placed in a different full scale category by the K-BIT than by the WISC-III.

SF4 vs. Full WISC-III

Twenty-seven participants completed both the Full WISC-III and

the WISC-III four subtest short form (SF4) in two separate sessions. The data from one participant were not included because the SF4 estimate of FSIQ minus WISC-III FSIQ for this subject (27 IQ points) was more than three standard scores away from the mean in the distribution of difference scores. Consequently this subject was considered an outlier and as such was likely to unduly influence the small data set. Of the remaining twenty-six participants, eleven had the SF4 in the first session followed by the full WISC-III in the second session, and fifteen had the full WISC-III in the first session followed by the SF4 in the second session. The mean retest interval was 17.15 days (range 4 - 36).

Table 3 shows summary IQ statistics for the full WISC-III and SF4 by administration order. SF4 IQ estimate and FSIQ obtained a significant correlation coefficient of $r = .91$, $p < .01$. The scores were analyzed as a repeated measures ANOVA comparing IQ results for the two tests and taking into account the order of administration. The administration order was the between-subjects factor, and the test (WISC-III vs. SF4) was the within-subjects factor. For FSIQ vs SF4 IQ estimate, the results showed a significant test effect ($F(1,24) = 22.68$, $p < .001$) in the direction of a SF4 overestimate (means = 94.58 versus

Table 3. Comparisons of SF4 and Full Scale IQ by Administrative Order.

	Short Form IQ[a]		Full Scale IQ	
Administration order	*M*	*SD*	*M*	*SD*
SF4 - Full WISC-III (n = 11)	98.09	11.11	95.18	11.17
Full WISC-III - SF4 (n = 15)	100.13	12.52	94.13	12.60

[a] Picture Completion, Information, Block Design and Vocabulary; see Sattler (1992).

99.27). The SF4 estimates were inclined to be higher when administered after the full scales (6 IQ points) versus before the full scale (3 IQ points; $F(1, 24) = 2.76$, $p = .11$). Overall, 22% of participants were placed in different Wechsler (1991) IQ categories by SF4 and 66% of SF4 IQ estimates were within the 95% confidence limits of FSIQ.

Test-Retest Reliability of the K-BIT

Twenty-two participants completed two administrations of the K-BIT in a standard test-retest methodology. The mean retest interval was 18.71 days (range 8 - 36). The data from one participant were not included in the analysis of the K-BIT Composite because the K-BIT minus retest K-BIT estimate for this subject (36 IQ points) was more than three standard scores away from the mean in the distribution of difference scores. Consequently this subject was considered an outlier and as such was likely to unduly influence the small data set. Descriptive statistics are presented in Table 4. Paired t-tests showed significant test-retest differences for Vocabulary ($t(21) = -4.98$, $p < .001$), Matrices ($t(21) = -2.10$, $p < .048$) and Composite IQ ($t(20) = -4.26$, $p < .001$). There was a four point practice effect for the Verbal, Matrices and Composite scales. Correlation Coefficients were $r = .91$, $p < .01$ for Verbal; $r = .69$, $p < .01$ for Matrices; and $r = .92$, $p < .01$ for Composite IQ.

Table 4. Comparisons of first and second administrations of the K-BIT.

	K-BIT			
	First Administration		Second Administration	
Subtest	*M*	*SD*	*M*	*SD*
Vocabulary (n = 22)	94.76	9.68	99.05	8.31
Matrices (n = 22)	91.52	15.09	95.19	14.08
Composite (n = 21)	92.43	11.74	96.81	10.76

Discussion

Results of this study show that the K-BIT does not perform well in terms of the validity criteria that are typically applied to short form estimates of intelligence. Although mean FSIQ and Performance IQ were not significantly different from K-BIT estimates, K-BIT Verbal IQ overestimated VIQ by an average of five points. Correlation Coefficients between the K-BIT and WISC-III were significant but not impressive (i.e., VIQ $\underline{r}$ = .73, PIQ $\underline{r}$ = .53, FSIQ $\underline{r}$ = .69). There was a 56% misclassification rate for the K-BIT in terms of FSIQ, and only 44% to 50% of the IQ estimates were within WISC-III 95% confidence limits. The four subtest short form performed reasonably well as an estimate of FSIQ. Although there was a mean overestimate of five IQ points, indicators are that this SF4 is a better estimate of FSIQ (3 points) when administered before the full scale. Similar results have been found in previous research (Thompson, 1995) where SF4 yielded a 3 point overestimate of FSIQ when administered first compared to an eleven point overestimate when SF4 was administered following the full scale. The correlation coefficient between SF4 and full scale WISC-III was high

and significant (r = .93), and only 22% of participants were placed in different IQ categories by SF4. Overall, 66% of SF4 IQ estimates were within the 95% confidence limits of WISC-III FSIQ. Clearly, SF4 is better than the K-BIT at estimating performance on the WISC-III. Although administration time was not measured, SF4 seemed to require about ten minutes more to complete. Presumably one advantage of the K-BIT is that it can provide estimates of VIQ, PIQ and FSIQ. However, according to the results of this study, K-BIT estimates of VIQ and PIQ have significant limitations. In previous studies VIQ and PIQ have either been underestimated (Kaufman & Kaufman, 1991; Prewett, 1992a; Prewett 1992b), or overestimated (Naugle et al., 1993; Prewett, 1992a) by the K-BIT. One study found no difference between WISC-R VIQ and K-BIT VIQ estimate for referred students, but the K-BIT underestimated PIQ in this case by over eight IQ points (Prewett, 1992b). While the research to date has shown that mean K-BIT estimates of FSIQ, VIQ and PIQ frequently over and underestimate, these results could be due to small sample sizes. A meta-analysis on the data from these studies would be useful to resolve the issue. For other validity criteria (i.e., correlations and misclassification rates) data have rarely been reported, or done so with inconsistency. Consequently it has been difficult to judge the

usefulness of the K-BIT as an individual screening instrument.

The results of this study also showed test-retest reliability results that were in keeping with coefficients reported in the K-BIT manual (test-retest intervals were 18.71 and 21.0 days respectively): Specifically results were Vocabulary $\underline{r}$ = .91 and .94, Matrices $\underline{r}$ = .69 and .85, and Composite $\underline{r}$ = .92 and .90, respectively. The K-BIT manual did not report the magnitude of observed practice effects. According to the present study, clinicians can expect about a four-point practice effect for Vocabulary, Matrices and Composite IQ scales when readministering the K-BIT.

When interpreting the results of this study it is important to keep in mind that data from one outlier were removed from each set of statistical analyses because its presence was expected to have an undue influence on the results. For the K-BIT, the outlier was 63 points different from the FSIQ. There were several other large differences between K-BIT and FSIQ that were included in the analysis (e.g., 3 differences over 25 points). For SF4, the outlier was 27 points different from FSIQ. There was one 25 point SF4-FSIQ difference which was included in the analysis. For K-BIT test-retest reliability, the outlier had a 36 point difference between K-BIT test and retest Composite IQ scores. There

was one difference over 25 points which was included in the analysis. While these individuals were easily identified through statistical means, clinically nothing suggested that the administration was invalid. These extreme score differences could be a product of the artificial situation of completing four intelligence tests within two sessions, differences in the participants' level of motivation between the two sessions, or may simply be a limitation of brief intelligence testing. At the same time, it may not be unusual for individuals in clinical settings to undergo this amount of testing, although not with four IQ measures. Regardless of their cause, it is important that clinicians be aware that differences between the brief intelligence test results and the WISC-III are possible. This is because there will always be subjects who happen to perform either much better, or much worse on the selected subtests than they do on others.

The test-retest methodology used in this study revealed a between subjects effect such that subjects who completed the K-BIT then WISC-III received higher scores on Verbal IQ than those who completed the WISC-III then K-BIT. While the Chi Square test of independence was not significant, it did reveal that 59% of participants who completed the K-BIT then WISC-III did so in two sessions, while 67% of participants who completed the WISC-III then K-BIT did so in one session. With a

small sample this distribution may in some way have effected the between-subjects effect.

Limitations of the current study include the small sample size (n = 35), and the likelihood that results from this sample of young offenders may not generalize to other clinical groups. While this study did not achieve a fully counterbalanced, randomized design, it is important for research to continue to approach this ideal. Data collection is continuing, so that this study may possibly be published with a larger sample size. The strengths of this project include the strong methodological design and careful checking of protocols for clerical and computational errors. Had this project not taken into account classification errors, hit rates for 95% confidence intervals, correlation coefficients as well as differences between the means, the full limitations of the short form estimates would not have been discovered. Hence, a comprehensive evaluation of short form validity using multiple criteria is recommended.

References

Adams, R. L., Smigielski, J., & Jenkins, R. L. (1984). Development of a Satz-Mogel short form of the WAIS-R. Journal of Consulting and Clinical Psychology, 52, 908.

Cyr, J. J. & Atkinson, L. (1991). Use of population - specific parameters in generating WAIS-R short forms. Psychological Reports, 69, 151-167.

Donders, J. (1992). Validity of two short forms of the WISC-R in children with traumatic brian injury. Journal of Clinical Psychology, 48, 364-370.

Dumont, R., & Faro, C. (1993). A WISC-III short form for learning-disabled students. Psychology in the Schools, 30, 212-218.

Fell, L., & Schmidt-Fell, S. (1982). Effectiveness of WISC-R short-forms in screening gifted children. Psychological Reports, 51, 1017-1018.

Haynes, J. P. (1983). Comparative validity of three Wechsler short forms for delinquents. Journal of Clinical Psychology, 39(2), 275-278.

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

Kaufman, A. S., & Kaufman, N. L. (1990). Kaufman Brief Intelligence Test. Manual. Circle Pines, MN: American Guidance Service.

Leckliter, I. N., Matarazzo, J. D., & Silverstein, A. B. (1986). A literature review of factor analytic studies of the WAIS-R. Journal of Clinical Psychology, 42, 332-342.

Matarazzo, J. D. (1972). Wechsler's measurement and appraisal of adult intelligence (5th ed.). Baltimore: Williams & Wilkins.

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

Plumridge, S. (1994). Two and four subtest short forms of the Wechsler Adult Intelligence Scale - Revised: A comparative validity study with normals. Unpublished Masters thesis, Lakehead University.

Prewett, P. N. (1992a). 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. (1992b). 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. (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.

Resnick, R. J. & Entin, A. D. (1971). Is an abbreviated form of the WISC valid for Afro-American children? Journal of Counselling and Clinical Psychology, 36, 97-99.

Ryan, J. J. (1981). Clinical utility of a WISC-R short form. Journal of Clinical Psychology, 37, 389-391.

Ryan, J. J. (1983). Clinical utility of a WAIS-R short form. Journal of Clinical Psychology, 39, 261-262.

Sattler, J. M. (1992). Assessment of Children: WISC-III and WPPSI-R supplement. San Diego, CA: Jerome M. Sattler, Publisher.

Satz, P., & Mogel, S. (1962). An abbreviation of the WAIS for clinical use. Journal of Clinical Psychology, 30, 97-99.

Silverstein, A. B. (1975). Validity of WISC-R short forms. Journal of Clinical Psychology, 31, 696-697.

Silverstein, A. B. (1982). Two- and four-subtest short forms of the Wechsler Adult Intelligence Scale - Revised. Journal of Consulting and Clinical Psychology, 50, 415-418.

Silverstein, A. B. (1985a). Two- and four-subtest short forms of the WAIS-R: A closer look at validity and reliability. Journal of Clinical Psychology, 41, 95-97.

Silverstein, A. B. (1985b). An appraisal of the three criteria for evaluating the usefulness of WAIS-R short forms. Journal of Clinical Psychology, 41, 676-680.

Silverstein, A. B. (1989). Agreement between a short form and the full scale as a function of the correlation between them. Journal of Clinical Psychology, 45, 929 - 931.

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

Thompson, A. P. (1987). Methodological issues in the clinical evaluation of two-and four-subtest short forms of the WAIS-R. Journal of Clinical Psychology, 43, 142-144.

Thompson, A. P. (1995). Test-retest evaluation of a four subtest WAIS-R short form with young offenders. Journal of Clinical Psychology, 51, 410-414.

Thompson, A. P., & LoBello, S. G. (1994). Reliable and abnormal scaled score range estimates for common Wechsler short forms. Journal of Psychoeducational Assessment, 12, 264-269.

Thompson, A. P., Howard, D., & Anderson, J. (1986). Two- and four-subtest short forms of the WAIS-R: Validity in a psychiatric sample. Canadian Journal of Behavioural Sciences, 18, 287-293.

Watkins, C. E. (1986). Validity and Usefulness of WAIS-R, WISC-R, and WPPSI short forms: A critical review. Professional Psychology: Research and Practice, 17, 36-43.

Watkins, C. C., Jr., McKay, B. L., Parra, R. A., & Polk, N. E. (1987). Using WAIS-R short forms with clinical outpatients: A cautionary note. Professional Psychology: Research and Practice, 17, 36-43.

Wechsler, D. (1974). WISC-R manual. Wechsler Intelligence Scale for Children - Revised. New York: Psychological Corporation.

Wechsler, D. (1981). WAIS-R manual. New York: Psychological Corporation.

Wechsler, D. (1989). WPPSI-R manual: Wechsler Preschool and Primary Scale of Intelligence - Revised. New York: Psychological Corporation.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children - third edition manual. New York: Psychological Corporation.

Appendix A

LAKEHEAD UNIVERSITY
5 Oliver Road, Thunder Bay, Ontario, Canada P7B 5E1
Office of the President
Telephone (807) 343-8200

7 March 1994

Ms. Janet Browne
Department of Psychology
Lakehead University
THUNDER BAY, ONTARIO
P7B 5E1

Dear Ms. Browne:

Based on the recommendation of the Ethics Advisory Committee, I am pleased to grant ethical approval to your research project entitled: TEST-RETEST AND VALIDITY ANALYSIS OF THE K-BIT AND SHORT FORM OF THE WISC-III.

Best wishes for a successful research project.

Sincerely,

ROBERT G. ROSEHART
President

/lw
cc: Dr. Tony Thompson

AGREEMENT

BETWEEN: *FRED SCHMIDT, PH.D.*
and
JANET BROWNE

AND: *WILLIAM W. CREIGHTON YOUTH SERVICES*

RE: RESEARCH PROJECT

This confirms approval for research conducted with clients of the Agency, under the following conditions:

a) The research relates to the proposal regarding "A Comparison of Short versus Long Intelligence Tests".

b) The researchers, Dr. Fred Schmidt and Janet Browne, are considered to be acting as representatives of William W. Creighton Youth Services, and sign Agency Oaths of Confidentiality.

c) Participation by clients is voluntary and subject to consent of both youth and their parent/guardian.

d) The Agency will provide access for the researchers to names, addresses and telephone numbers of parents/guardians to the researchers for purposes of obtaining consents.

e) The Agency may screen clients for referral to the project, taking into account security needs or youth's current disposition.

SIGNED:

Fred Schmidt — Date: July 4/94

Janet Browne — Date: 94.09.19

For the Agency — Date: July 4/94

Appendix B

LAKEHEAD UNIVERSITY

; Oliver Road, Thunder Bay, Ontario, Canada P7B 5E1

Department of Psychology
Telephone (807) 343-8441

A COMPARISON OF SHORT VERSUS LONG INTELLIGENCE TESTS

DEAR :

I am asking you to participate in a research project, to be carried out at Creighton Youth Services. The project aims to determine the usefulness of a new short intelligence test (30 minutes) in comparison with the longer standard intelligence test (60 minutes) currently in use. The research is being conducted by myself, Janet Browne, Masters Student, Lakehead University, under the supervision of Dr. Anthony P. Thompson, Department of Psychology, Lakehead University, and Dr. Fred Schmidt, Registered Psychologist, Lakehead Regional Family Centre.

You will be asked to participate in two intelligence testing sessions over the next two months. One session will last about an hour, the other about an hour and a half. Tests will be administered individually and in private. Your participation in this study is entirely voluntary. You will be at no disadvantage on the YO unit if you do not participate and there are no special considerations on the YO unit for those who do participate. We do not expect that you will be upset by the testing procedure. Most people enjoy the tasks, although some people are anxious because they want to do well. If you decide to participate, you may withdraw from the testing at any time.

Results of the tests will be kept confidential, and individual scores will not be released to Creighton Youth Services. Your parents also have the right to the same test feedback that you receive. However, we would discuss this with you prior to acting upon such a request. We are interested in group results rather than individual scores. You will be able to receive some feedback concerning your own performance once testing has been completed, and we will talk with you about any concerns you may have about your results. An effort will be made to explain "intelligence" to you. The tests used in this study measure only some of the ways in which people can be intelligent.

Your participation in this study will reveal new and useful information concerning the process of intellectual assessment. Results will only be used for research purposes, and your name will not appear in reports or publications. If you agree to participate in this project, please complete the attached consent form. If you have any questions or concerns in relation to this research project, please feel free to contact either myself at 343-8476 or Dr. A. P. Thompson at 343-8646. Ethical approval for this research has been received from the Lakehead University Ethics Committee, Lakehead Regional Family Centre, and Creighton Youth Services.

Yours Truly,

Janet Browne,
Department of Psychology, Lakehead University.

LAKEHEAD UNIVERSITY

5 Oliver Road, Thunder Bay, Ontario, Canada P7B 5E1

Department of Psychology
Telephone (807) 343-8441

A COMPARISON OF SHORT VERSUS LONG INTELLIGENCE TESTS

DEAR PARENT/GUARDIAN:

I am asking for your permission to have your son/daughter participate in a research project, to be carried out at the Creighton Youth Services. The project aims to determine the usefulness of a new short intelligence test (30 minutes) as compared to a longer intelligence test (60 minutes) which is frequently used. The research is being conducted by myself, Janet Browne, Masters Student, Lakehead University, under the supervision of Dr. Anthony P. Thompson, Department of Psychology, Lakehead University, and Dr. Fred Schmidt, Registered Psychologist, Lakehead Regional Family Centre.

Your child will be asked to participate in two intelligence testing sessions over the next two months. One session will last about one hour, the other about an hour and a half. Tests will be administered individually and in private. Your child's participation in this study is completely voluntary. There are no disadvantages on the YO unit if your child does not participate, and there are no special considerations on the YO unit for those who do participate. If your child does participate, they may withdraw from the testing at any time. Results will only be used for research purposes, and individual participants will not be identified in reports or publications. Your child's participation in this study will reveal new and useful information concerning the process of intellectual assessment.

Results of the tests will be kept confidential, and individual scores will not be released to Creighton Youth Services. We are interested in group results rather than individual scores. However, your child will be able to receive some feedback concerning their own performance once testing has been completed. We do not expect that they will be upset by the testing procedure. Most people enjoy the tasks, although some people are anxious because they want to do well. An effort will be made to explain the concept of "intelligence" to your child, and to discuss any concerns they may have about their results. The tests used in this study measure only some of the ways in which people can be intelligent.

If you approve of your child's participation in this project, please complete the attached consent form, and return it to Creighton Youth Services. If you have any questions or concerns in relation to this research project, please feel free to contact either myself at 343-8476 or Dr. A. P. Thompson at 343-8646. Ethical approval for this research has been received from the Lakehead University Ethics Committee, Lakehead Regional Family Centre, and Creighton Youth Services.

Yours Truly,

Janet Browne,
Department of Psychology, Lakehead University.

Appendix C

LAKEHEAD UNIVERSITY

5 Oliver Road, Thunder Bay, Ontario, Canada P7B 5E1

Department of Psychology
Telephone (807) 343-8441

PARTICIPANT CONSENT FORM

I, _____ agree to participate in the study
(Full Name)
comparing a short and a long intelligence test. The study will be conducted by Janet Browne, Masters Student, Lakehead University (343-8476), Dr. Anthony P. Thompson, Department of Psychology, Lakehead University (343-8646), and Dr. Fred Schmidt, Registered Psychologist, Lakehead Regional Family Centre.

I understand that I will be given an intelligence test on two separate occasions. One session will last about an hour, the other about an hour and a half. The risks and benefits of participation have been explained to me.

I understand that all information will be kept confidential, and that I may withdraw from this research project at any time without penalty. I will obtain some feedback on my performance after the second testing session, and may also obtain a brief account of the results of the study.

_____ _____
Signature Date

_____ _____
Witness Date

LAKEHEAD UNIVERSITY

5 Oliver Road, Thunder Bay, Ontario, Canada P7B 5E1

Department of Psychology
Telephone (807) 343-8441

PARENT/GUARDIAN CONSENT FORM

I, _____ agree to allow my son/daughter,

_____ to participate in the study comparing
(Full Name)
a short and a long intelligence testing. The study will be conducted by Janet Browne, Masters Student, Lakehead University (343-8476), under the supervision of Dr. Anthony P. Thompson, Department of Psychology, Lakehead University (343-8646), and Dr. Fred Schmidt, Registered Psychologist, Lakehead Regional Family Centre.

I understand that my son or daughter's participation will entail being given an intelligence test on two separate occasions, and that he/she will receive some feedback. The risks and benefits have been explained to me. I also understand that my son/daughter may withdraw at any time.

_____ _____

Signature Date

Appendix D

Feedback Protocol

I wanted to give you feedback on the first test that you took. First, I would like to define "intelligence" for you. Basically, intelligence is your ability to solve problems. The long intelligence test that you took breaks this ability down into two areas. The first is Verbal Intelligence. Verbal Intelligence is your ability to solve problems using words and numbers. The second is Performance Intelligence. Performance Intelligence is your ability to solve problems visually, often by manipulating objects with your hands, like the puzzles that you did.

Your particular results were: (feedback given only on the Full WISC-III, using 95% confidence intervals and including a range only if the score fell into it by at least 3 IQ points.)

1. Your Verbal Intelligence was in the (score given according to the Wechsler classification) range.
2. Your performance, or non-verbal intelligence was in the _____ range.
3. When your put these two results together, your overall problem solving ability as measured by the test you took was in the _____ range.

Do you have any questions? Would you like anything repeated?

This is not the absolute or final word on your intellectual level. There are three things that I'd like you to keep in mind. First, the conditions under which you took the test can influence the results. For example, if your were nervous, distracted, or just not trying. Secondly, I would like to remind you that intelligence is only one factor related to success. You also need motivation and effort. So, people with high intelligence can squander their ability, and people with lower intelligence can be successful with persistence. Lastly, the test you took doesn't measure all aspects of intelligence, such as musical aptitude, athletic ability, social skills or aptitude to succeed in business.

Just to make sure that you understand all that I have just told you, I'd like for you to repeat back to me what you have learned. Do you have any more questions?