COMPARABILITY OF THE WPPSI-R AND THE STANFORD-BINET: FOURTH EDITION

THESIS

Presented to the Graduate Council of the University of North Texas in Partial Fulfillment of the Requirements For the Degree of

Master of Science

By

Catherine Bass, B.A.

Denton, Texas

May, 1990

The purpose of this study was to compare the performance of children on the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) with their performance on the Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE). One hundred and four children between 3 and 7 years of age were administered both tests. A moderate correlation was found between the WPPSI-R Full Scale IQ and the SB:FE Composite Score with a Pearson product-moment correlation of .46. This correlation suggests that the two tests are not interchangeable measures of children’s intelligence. They may measure different, equally important aspects of intelligence. As both tests used are relatively new, the current findings should be considered one step in the accumulation of knowledge about the usefulness of the WPPSI-R.
TABLE OF CONTENTS

LIST OF TABLES ........................................ iv

COMPARABILITY OF THE WPPSI-R AND THE STANFORD-
BINET: FOURTH EDITION

Introduction ........................................ 1

   Development of the WPPSI
   Description of WPPSI Subtests
   Reliability of the WPPSI
   Validity of the WPPSI
   Development of the WPPSI-R
   Standardization of the WPPSI-R
   Limitations of the WPPSI
   Changes from the WPPSI
   Reliability of the WPPSI-R
   Validity of the WPPSI-R
   Stanford-Binet: Fourth Edition
   Limitations of the Stanford-Binet: Fourth
   Edition
   Summary and Purpose of the Study

Method ............................................. 26

   Participants
   Measures
   Examiners
   Procedures

Results ............................................. 27

Discussion ......................................... 30

APPENDICES ......................................... 34

REFERENCES ......................................... 42
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
</table>
A commonly used measure for assessing the intellectual level of young children is the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) (Wechsler, 1967). This is one of three Wechsler scales currently in use. The others are the Wechsler Adult Intelligence Scale - Revised (WAIS-R) (Wechsler, 1981), and the Wechsler Intelligence Scale for Children - Revised (WISC-R) (Wechsler, 1974). In the last decade, both the WAIS and the WISC were revised to update the test items and the normative data. Recently, the WPSSI has been similarly revised.

The revision of any intelligence test requires the examination of the resulting psychometric properties. One aspect of this examination consists of exploring the new test's validity. One way in which validity may be estimated is by comparing scores on the new measure with scores on existing measures. The current study compares scores obtained on the WPPSI-R to those obtained concurrently on the Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE) (Thorndike, Hagen, & Sattler, 1986). This is the latest edition of the Stanford-Binet scales first published in 1916 (Binet & Simon, 1916).
**Development of the WPPSI**

Wechsler described intelligence as "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment," and designed his tests to yield an overall IQ representing an index of general mental ability (Sattler, 1988). All Wechsler tests include multiple subtests assumed to measure different aspects of intelligence. Three deviation IQ's are obtained from Wechsler tests. These are a Verbal, Performance, and Full Scale IQ, each obtained by comparing scores with those of the normative group of the same age.

In developing his first intelligence scale, the Wechsler-Bellevue Intelligence Scale, Form I, Wechsler adapted several standardized tests available in the late 1930's. He used the Army Alpha for Information and Comprehension subtests, the Stanford-Binet for Comprehension, Arithmetic, Digit Span, Similarities, and Vocabulary subtests, the Healy Picture Completion Tests and other tests with picture items for Picture Completion and Picture Arrangement subtests, the Kohs Block Design Test for a Block Design subtest, and the Army Beta for Digit Symbol and Coding subtests.

The WPPSI was published in 1967 following publication of the WAIS and the WISC. Although similar to the WISC
(and subsequently the WISC-R), the WPPSI is not considered to be a downward extension of this test. Eight of its eleven subtests are adapted from WISC subtests and three are unique to the WPPSI.

Description of WPPSI Subtests

Verbal subtests. The Information subtest requires the child to answer brief oral questions demonstrating his or her knowledge about objects or events in the environment. Vocabulary requires verbal definitions for orally presented words. Arithmetic requires a demonstration of basic quantitative concepts through picture items, simple counting tasks, and more difficult word problems. In the Similarities subtest, an understanding of the concepts of similarity is demonstrated in two ways. First, the child completes a verbally presented sentence reflecting a similarity. A second section requires the child to explain the similarity between two objects or events. Comprehension requires the expression in words of a child's understanding of reasons for actions and consequences of events. Sentences is an optional subtest designed to be similar to the Digit Span subtest on the WISC, using meaningful verbal stimuli instead of numbers. The child is asked to repeat sentences verbatim.

Performance subtests. Animal House involves the placement of different colored pegs in their appropriate
roles corresponding to a model. Picture Completion requires the child to identify the missing element in pictures of common objects or events. The Mazes subtest involves the solving of pencil and paper mazes under time constraints. Geometric Design involves reproducing a geometric figure from a printed model. Finally, the Block Design subtest requires the analysis and reproduction of patterns made from flat, two-colored blocks within a specified time limit.

Standardization of the WPPSI

The original WPPSI standardization sample consisted of 1,200 children including 100 girls and 100 boys in each of six age groups between 4 and 6 1/2 years. The sample is said by the publishers to reflect the 1960 U.S. census data (Wechsler, 1967).

Reliability of the WPPSI

Since the WPPSI-R is designed to replace the WPPSI, a review of the research examining the original test was important as a basis for the evaluation of the anticipated psychometric properties of the revision. Reliability coefficients for WPPSI subtests were found by the split-half method using the Spearman-Brown formula. Animal House scores depend largely on speed, therefore the reliability of this subtest was found by the test-retest method. Reliability coefficients were found for each half-year age
group in the standardization sample, with most of the resulting coefficients falling in the .80 to .90 range. Reliabilities ranged from .80 to .90 for the Verbal IQ, from .84 to .91 for the Performance IQ, and from .92 to .94 for the Full Scale IQ (Anastasi, 1982).

Sattler (1988) reports that the average internal consistency coefficients are .96 for the Full Scale IQ, .94 for the Verbal IQ, and .93 for the Performance IQ. Test-retest reliability has been found to range from .86 to .91 for the three scales, with the performance scale subtests showing greater practice effect than the Verbal scale subtests. Wasik and Wasik (1970) readministered the WPPSI to 50 children in the standardization sample after approximately 11 weeks. They found increases on the Verbal, Performance, and Full Scale IQs of 3.0, 6.6, and 3.6, respectively. Tew and Laurence (1983) administered the WPPSI at 5 years and again at 6 1/2 years of age to 51 Welsh children with spina bifida. They found correlations of .88 (Verbal Scale), .90 (Performance Scale), and .92 (Full Scale). Mean IQ differences were less than 6 points, indicating good stability in the WPPSI.

Subtest reliabilities are less satisfactory than those of the three scales. Average subtest reliabilities range from .77 for Animal House to .87 for Mazes and they are similar for Verbal and Performance subtests. These
differences must be considered in interpreting test profiles as some subtests' reliability may be less adequate at younger ages (Sattler, 1988). These reliability results were approximately the same for the WPPSI-R.

The WPPSI has been shown to have good reliability with special populations. Researchers have reported split-half reliabilities of .95 for Mexican-American children (Henderson & Rankin, 1973), .93 for gifted children (Ruschival & Way, 1971), .89 for black, lower socioeconomic children (Croake, Keller, & Catlin, 1973), and .88 for mentally retarded children (Richards, 1970).

Validity of the WPPSI

In investigating the psychometric properties of the WPPSI-R, it is useful to consider the comparability of the original version with other tests designed to measure children's intelligence. Although no section specifically addressing validity is included in the WPPSI manual, it does provide some information relative to the instrument's validity. Subtest intercorrelations within each age level in the standardization sample range largely between .40 and .60. Correlations between Verbal and Performance subtests approach those within each scale. The average correlation between Verbal and Performance IQs is .66 (Anastasi, 1982).

The theoretical validity of the WPPSI has been explored. Silverstein (1986) analyzed the intercorrelations
among the eleven subtests of the WPPSI for each of the six age groups in the standardization sample. Clusters were found that clearly corresponded to the Verbal and Performance scales. Results were in general agreement with those of previous research in which factor analysis was employed with the same data. Sattler (1988) reports the WPPSI was correlated with the 1960 norms of the Stanford-Binet and was found to have satisfactory concurrent validity. Carlson (1981) factor analyzed the WPPSI at each of six age levels between 4 and 6 1/2 years through the method of principal factors. A two-factor solution corresponding essentially to Wechsler’s (1967) a priori grouping of the subtests into the Verbal and Performance scales best described the eleven WPPSI subtests. Examination of the common, specific, and error variance components of the subtests across age levels indicated adequate subtest specificity to allow for individual interpretation of all subtests except Information and Comprehension. Ramanaiah (1979) completed a study investigating Wechsler’s hypothesis that verbal and performance factors are the major dimensions underlying the WISC and the WPPSI subtests. A confirmatory factor analysis strategy was used. In general, correlations between the WPPSI and the WISC-R were especially high with a range of .73 to .86 for the Verbal, Performance, and Full Scales.
A difficulty in interpreting WPPSI scores relates to factor-analytic studies. While most studies support the general, verbal and performance factors in the test, the latter two are less clearly-delineated in younger children aged 4 to 4-1/2 years. These differences, related to the organization of abilities in younger children, result in difficulties when interpreting these children’s test results. An additional problem with factor-analytic studies for all age groups is their failure to include more behavioral indices external to the scales themselves (Anastasi, 1982). These same problems continue in the WPPSI-R.

Kaufman (1973) reanalyzed the WPPSI standardization sample to obtain information about the relation of WPPSI scores to socioeconomic status, urban versus rural residence, and geographic region. Socioeconomic status (as defined by father’s occupation) yielded significant differences, while the other two variables did not. Children whose fathers worked in the professional and technical occupations scored significantly higher than other groups (mean IQ = 110), and children whose fathers worked in unskilled occupations scored significantly lower (mean IQ = 92.1) (Anastasi, 1982). This needs to be remembered in interpreting WPPSI IQs, as scores may be
affected by cultural and educational opportunities and attitudes rather than general intelligence.

Yule (1982) conducted longitudinal studies in which 85 children were tested using the WPPSI at 5 1/2 years and the WISC-R at 16 1/2 years. The resulting correlations were .82, .73, and .86 for the Verbal, Performance and Full Scale IQ's, respectively. The mean WPPSI IQs were 103.26, 107.97, and 106.01 for the respective scales. The subsequent mean WISC-R scores were 99.44, 104.78, and 102.78 for the three scales, indicating the relatively good predictive validity of the WPPSI.

The concurrent validity of the WPPSI has also been examined. Reynolds (1981) compared correlations of WISC-R and WPPSI IQs with subtest scores on the Wide Range Achievement Test (WRAT) of 110 five to 11 year olds referred by regular classroom teachers for school psychological services. No significant differences in the magnitude of the correlations of the WPPSI and the WISC-R with the WRAT-R were found. Results support the use of the two instruments as equivalent predictors of achievement.

Another study which supports the validity of the WPPSI with other measures of intelligence was completed by Arinoldo (1982). He compared scores on the WPPSI, WISC-R and McCarthy Scales of Children's Abilities. Participants were 40 four to five and one-half and seven to eight and one-
half year old children. Moderate to strong correlations were found between the General Cognitive Index of the McCarthy Scales and Wechsler's IQs at the preschool and school-age levels. Another study of the WPPSI and the McCarthy scales was conducted by Schmitz and Beckenbaugh (1979) using 21 predominantly white, middle class children between three years, ten months of age and five years, seven months of age. The resulting correlation was .86, with the mean WPPSI Full Scale IQ being seven points higher than the mean McCarthy Scale General Cognitive Index.

Finally, Field (1987) found that the WPPSI Full Scale IQ correlated significantly with scores on the Stanford-Binet Form L-M and an adaptation of the Leiter International Performance Scale for language-impaired and/or behaviorally disturbed preschool children. In this study, the WPPSI was administered an average of 16 months after the other measures. Participants were 26 children on a hospital child development unit, chosen because they had at least a 16 point difference between their Leiter and Stanford-Binet scores. The children had a mean age of 48.15 months, a mean Leiter score of 102.96, and a mean Stanford-Binet score of 67.69. The subsequent WPPSI scores were lower than the Leiter scores and higher than the Stanford-Binet scores.
Limitations of the WPPSI

There are several limitations related to the interpretation of the WPPSI. Administration time may be problematic for some children. Ten percent of the standardization sample required 90 minutes or more to complete the test and administrations lasting over an hour are common. This often results in two testing sessions and little empirical data exists concerning the effect of breaks on test scores. Administration time is unfortunately longer for the WPPSI-R.

Another limitation of the test is its limited floor and ceiling. For four-year-old children, the lowest Verbal and Performance Scale IQs are 55 and the lowest Full Scale IQ is 51. Children receive as much as 4 points for no correct responses. The usefulness of the test is therefore questionable for younger children and for children functioning at greater than 2 standard deviations below the mean of the scale (Sattler, 1988). Similarly, the limited ceiling limits the tests' usefulness for intellectually gifted children.

As with other Wechsler tests, the scoring of certain subtests is difficult and subjective. This is especially true of the Geometric Design, Vocabulary, Similarities and Comprehension subtests. Consultation with other
professionals is advised when responses are questionable. WPPSI-R scoring has improved on these issues.

**Development of the WPPSI-R**

After 20 years of use, the WPPSI has been revised to update the norms and to extend the age range of the scale downward from 4 to 3 years and upward from 6 1/2 to 7 years. While none of the original subtests have been eliminated, some subtests have been revised and others added. Changes within subtests include (a) eliminating items that are also included in the WISC-R, (b) eliminating questions considered to be obsolete or biased, and (c) adding a section of simple introductory items at the beginning of several subtests to help introduce the test and prepare children for the task. An Object Assembly subtest similar to that in the WISC-R and WAIS-R was added. The addition of items and subtests resulted in the extension of administration time from 50-75 minutes to 90-120 minutes (Psychological Corporation, 1987).

In revising the WPPSI, the publishers initially incorporated changes from two sources (Wechsler, 1989). First, a panel of experts and users in the area of intellectual assessment examined extensively the content and format of the WPPSI subtests. The result of this examination was changes in the test’s artwork, materials, and instructions. Second, new items were created to extend
the age range downward and upward. Where suitable items could not be created by modifying existing items, new items were developed. These include pictures for the initial items on the Vocabulary subtest and visual recognition/discrimination items on the Geometric Design subtest.

These modifications were incorporated into a pilot version of the WPPSI-R, which was used in several testings to (1) indicate which items performed as intended and which did not, and (2) to determine whether the directions were clear and understandable to children.

After changes were made based on these pilot studies, the new version was evaluated by administering it to over 500 children nationally at three age levels, resulting in deletion of more items. Then, a panel of reviewers evaluated each item in this edition for racial and sex bias. These items were revised or eliminated. The standardization edition was then developed based on the empirical data and the reviewers' evaluations. Again, items found to have poor discriminating power, to be difficult to administer or score, or those found to be biased were eliminated from the final edition.

Standardization of the WPPSI-R

The standardization sample consists of 1,700 children between the ages of three years and seven years, 3 months. The sample is considered to be representative of the U.S.
population according to March 1986 data gathered by the Bureau of the Census. Six variables were used to determine the selection of the sample: age, sex, geographic region, ethnicity, parental education, and parental occupation.

The sample consisted of 100 boys and 100 girls in each of eight age groups, ranging in half years from three to seven years of age. Another group of 50 boys and 50 girls was used, ranging in age from seven years to seven years, 3 months. The sample represented the four major geographic regions of the United States: Northeast, North Central, South, and West, in accordance with the 1986 census. Proportions of Whites, Blacks, Hispanics and other ethnic groups were also represented in accordance with this census. Socioeconomic status also reflected this census and was determined by two variables: the education level and occupation of the parents. This is in contrast to the WPPSI standardization, which used urban-rural residence in place of parental education as a stratification variable.

Changes from the WPPSI

The WPPSI-R includes approximately 50 percent of the WPPSI items (excluding the Animal Pegs subtest) either unchanged or with slight modifications. The Performance and Verbal subtests are presented in alternating order as they were on the WPPSI. The revision, however, begins
with a new Object Assembly subtest, as most children find this test interesting. The WPPSI began with the Information subtest. Specific changes in test content, administration, and scoring are described below.

**Object Assembly.** This subtest, patterned after Object Assembly subtests on other Wechsler scales, is the only subtest to be added to the new test. Unlike other scales, these puzzles are printed in full color. The scoring has been modified to give credit both for correct assembly and speed.

**Geometric Design.** This subtest now has two sections, beginning with seven visual recognition-discrimination items. In the second part, three new designs have been added and four of the original WPPSI designs deleted. In scoring, specific criteria for each design are listed in order to make the process more objective. An additional change concerns second attempts. In the WPPSI, the better attempt was scored, whereas the second attempt is scored on the revision.

**Block Design.** Easier items have been added to the beginning of this subtest and more difficult items have been added to the end in order to extend the range. The only change in administration and scoring is that bonus points may now be earned for speed and accuracy.
Mazes. To accommodate the younger child, several new horizontal mazes have been added. The original chick and hen have been replaced by a squirrel and tree, and the stick figures were replaced with silhouettes of children. The mazes are now in booklet form rather than individual pages. Several demonstrations are now required with younger children along with verbal encouragement and cautions. Older children now begin at item 3 and are given additional instructions if this maze is failed. Crossing walls is now defined as an error, unlike in the original test.

Picture Completion. Several items have been added to this subtest and the artwork along with several items have been changed to add appeal. Younger children have been accommodated by this subtest with the addition of several easier items at the beginning. Additionally, children begin the test with an unscored sample and are given feedback on their performance on the first two items. Children aged five and over are administered this sample only if their initial items are failed.

Animal Pegs. The name of this subtest has been changed from Animal House and the artwork has been updated. Significantly, this subtest is no longer included in the Performance IQ, but is now optional. Although administration and scoring are unchanged, there is no optional retest.
Information. Several full-color picture items have been added to this subtest to extend the age range downward. In addition, five new verbal items have been added. Scoring criteria have been expanded with a wider range of acceptable responses.

Comprehension. Two-thirds of the items in this subtest were replaced to update the test content. Children are now given the two-point response if needed on items 1 and 2, whereas no help was admissible on the WPPSI. The number of items requiring two correct responses has been reduced.

Arithmetic. Three picture items were added to the beginning of this subtest to measure younger children's knowledge of quantitative relationships. The content of other items is essentially unchanged, although full-color art has been added to increase appeal. Administration and scoring are very similar to those of the WPPSI, with minor changes in establishing a basal level.

Vocabulary. Three basal items have been added which require the child to identify pictured objects. Additionally, seven verbal definition items have been added. Scoring has been made easier with more examples of 1- and 2-point responses.

Similarities. This subtest now consists of three sections. Part 1 includes picture items, Part 2 contains
sentence completion items, and Part 3 consists of verbal analogies. Most of the items in Part 3 are new. Scoring procedures are largely unchanged, with one point given for correct picture and sentence completion items and one or two points given for each correct verbal analogy item.

Sentences. Still an optional subtest, five sentences have been added to extend it downward and upward. The various types of errors made can now be recorded on the test record.

Reliability of the WPPSI-R

The WPPSI-R Manual (Wechsler, 1989) contains the only information available on the reliability of the test. Split-half reliabilities were computed by examining the items in each subtest, pairing them according to their level of difficulty and discriminating power, then splitting these pairs into alternate halves. For each subtest, raw scores of the two halves were correlated with the resulting coefficient corrected by the Spearman-Brown formula. This procedure was conducted for each of nine age groups, ranging in half-years from three to seven years of age. Each group contained 200 children, except the 7-year-old group, which contained 100 children.

The resulting correlations may be reviewed in Table 1. The split-half reliability of the Performance IQ ranged from .85 at 7 years to .93 at 3 to 4-1/2 years, with an average
of .92. The reliability of the Verbal IQ ranged from .86 at 7 years to .96 at 3 to 4-1/2 years, with an average reliability of .95. The reliability of the Full Scale IQ ranged from .90 at 7 years to .97 for 3 to 4-1/2 years, with an average reliability of .96. Individual subtest reliabilities ranged from .54 for Object Assembly at 7 years to .93 for Picture Completion at 4-1/2 years. Average split-half reliabilities across age groups for Performance subtests were: .63 for Object Assembly, .79 for Geometric Design, .85 for Block Design, .77 for Mazes, and .85 for Picture Completion. No split-half reliability is reported for Animal Pegs, as this method is inappropriate for tests of speed.

The stability of the WPPSI-R scores over time was assessed in another study cited in the manual. A random sample of 175 children was chosen from the standardization sample. Two age groups were formed: 36 to 59 months and 60 to 87 months. The testing interval ranged from 3 to 7 weeks with an average of 4 weeks. The resulting uncorrected test-retest reliability coefficients were .87 for the Performance IQ, .89 for the Verbal IQ, and .91 for the Full Scale IQ. Reliabilities for the individual Performance subtests were: .53 for Object Assembly, .66 for Geometric Design, .77 for Block Design, .55 for Mazes, .81 for Picture Completion, and .58 for Animal Pegs. The Verbal subtest reliabilities were:
.78 for Information, .77 for Comprehension, .71 for Arithmetic, .71 for Vocabulary, .73 for Similarities, and .76 for Sentences.

Interscorer reliability of the WPPSI-R was computed for the subtests which involve subjectivity in scoring. These are Comprehension, Vocabulary, Similarities, and Geometric Design. For this study, also cited in the manual (Wechsler, 1989), a sample of 151 children (83 males and 68 females) was randomly selected from the standardization sample. A group of scorers was trained and two scorers were randomly selected to score all the tests in each of several age groups. The resulting interscorer reliability coefficients were as follows: .96 for Comprehension, .94 for Vocabulary, .96 for Similarities, .94 for Mazes, and .88 for Geometric Design.

In summary, the reliability coefficients of the WPPSI-R are adequate and compare favorably with those of other children's intelligence tests.

Validity of the WPPSI-R

The WPPSI Manual (Wechsler, 1989) reports several studies comparing the test with other children's intelligence measures. In a study of the WPPSI and the WPPSI-R, 144 children were randomly selected from the standardization sample. The sample consisted of 73 females and 71 males between the ages of 48 and 49 months and is
reported to be consistent with the 1986 census data. Sixty-one children received the WPPSI-R first and 83 received the WPPSI first. The tests were administered from 3 to 5 weeks apart. The resulting correlations between the WPPSI-R and WPPSI Performance, Verbal, and Full Scale IQs were .82, .85, and .87, respectively, indicating good concurrent validity.

The WPPSI-R IQ scores were lower than those of the WPPSI by 9, 5, and 8 points for the Performance, Verbal, and Full Scale IQs, respectively. This is explained in the manual by noting the 22 year difference between standardization dates.

The WPPSI-R was compared to the WISC-R in another study conducted by Urbina and Clayton (cited in Wechsler, 1989). Twenty-five males and 25 females ranging from 72 to 86 months of age were administered both tests. Half of the children received the WPPSI-R first and half received the WISC-R first. The interval between tests ranged from 7 to 38 days, with an average of 19 days. Correlations between the two tests' Performance, Verbal and Full Scale IQs were .75, .76, and .85, respectively. Again, the WISC-R produced higher scores than the WPPSI-R by 9, 5, and 7 points on the Performance, Verbal, and Full Scale IQs, respectively.
Another study compared the WPPSI-R to the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983). Fifty-nine children were tested, ranging in age from 37 to 76 months (mean = 61 months). Correlations between the K-ABC Mental Processing Composite and the WPPSI-R Performance, Verbal, and Full Scale IQs were .41, .42, and .49, respectively. The mean WPPSI-R Full Scale IQ was 6 points lower than the mean K-ABC Mental Processing Composite. These moderate correlations and differences in means may indicate that the two scales measure different constructs.

The WPPSI-R was compared to the McCarthy Scales of Children’s Abilities (MSCA) in a sample of 93 children. The sample included 44 males and 49 females aged 4 to 6 years (mean = 62.5 months). The WPPSI-R Full Scale IQ and MSCA General Cognitive Index (GCI) correlated at .81. The difference of 2 points between the Full Scale IQ and GCI indicates that the two scales produce similar scores for children 4 to 6 years of age.

Stanford-Binet: Fourth Edition

The Stanford-Binet Intelligence Test is one of the most widely-used measures of intelligence and is therefore important in the determination of the validity of the WPPSI-R. Because the SB:FE is relatively new, however, relatively little research has been published regarding its
psychometric properties. The current version was published in 1986. This test consists of 15 subtests which are divided into four areas. Eight of these subtests are included in the battery for four to seven year-olds. The Verbal Reasoning Area for this age group consists of Vocabulary, Comprehension, and Absurdities subtests. The Abstract/Visual Reasoning Area includes Pattern Analysis and Copying subtests. The Quantitative Reasoning Area for this age group includes only one subtest by the same name. The Short-term Memory Area consists of Bead Memory and Memory for Sentences.

Raw scores from the subtests are converted into (a) subtest scaled scores with a mean of 50 and a standard deviation of 8, (b) Area scores with a mean of 100 and a standard deviation of 16, and (c) a Composite score with a mean of 100 and a standard deviation of 16. The Composite score is considered to be comparable to the deviation IQ's of the Wechsler scales. While the standard deviation of the Wechsler scales is 15, that of the SB:FE is 16, which is consistent with previous Stanford-Binet editions.

The standardization sample includes 5,013 people in 17 age groups from two years, zero months to 23 years, 11 months. The manual (Thorndike et al., 1986) states that the sample reflects the 1980 U.S. census data. Although variables such as geographic region, community size, ethnic
group, age, gender, and socioeconomic status were considered, weighting procedures were necessary to compensate for the inclusion of too many high SES children in the sample.

The reliability of the SB:FE Composite score is reported in the manual to be high (Thorndike et al., 1986). The median reliability coefficient is reported to be .97, with internal consistency reliabilities ranging from .73 to .94. Test-retest reliability was determined by testing two groups after a two to eight month interval. One group of 57 children had a mean age of five years, two months and the other group of 55 children had a mean age of eight years, one month. In the younger group, the test-retest reliability coefficient for the Composite score was .91 (Thorndike et al., 1986). Reliability coefficients for the subtests ranged from .56 to .78.

The manual also includes 13 studies on the validity of the SB:FE. These studies yield correlations between .27 and .91, with a median correlation of .80. The SB:FE yielded lower scores than the criterion test in 10 of the 13 studies. Other concurrent validity studies indicate that the SB:FE has adequate concurrent validity for populations within the average range of intelligence, but may yield lower scores than the other tests (Thorndike et al., 1986).
**Limitations of the Stanford Binet: Fourth Edition**

There are some limitations that should be remembered when interpreting the SB:FE scores. They include: (1) the lack of a comparable battery throughout the age ranges covered, (2) the variable range of scores at different ages, (3) the lack of support for the four area scores, and (4) the incorrect, overly high entry level points indicated in the manual, sometimes resulting in a prolonged administration due to beginning the test at too difficult a level (Sattler, 1988).

**Summary and Purpose of the Study**

Recently, the WPPSI has been revised in order to update test items, reestablish norms, and extend the age range downward and upward. When a new scale is developed or an existing measure is revised, research exploring its psychometric properties and the extent to which it correlates with existing measures is essential.

Previous research has shown that, although problems exist with the WPPSI, it correlates highly with some measures of intelligence as well as various measures of achievement. Findings from such research are of value in order to compare an individual's or a group's performance from one measure to another. Thus the purpose of this study was to compare the performance of children on the WPPSI-R with their performance on the SB:FE.
Method

Participants

Participants were 104 children from four different educational institutions in the suburbs of a large southern city; two public schools, one private school and one preschool. Forty-one males and 63 females participated, ranging in age from 3.1 to 6.9 years, with an average age of 6.2 years. Participants were selected by sending information about the study and a consent form to all parents in each school who had children four to seven years of age. Children whose parents returned the consent form were selected for the study.

Measures

Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R). The standardization version of this test, provided by the Psychological Corporation, was used in this study. This instrument has been described in an earlier section of this manuscript.

Stanford-Binet Intelligence Scale: Fourth Edition. This measure has been described previously.

Background Information Questionnaire. This questionnaire consists of nineteen items which were developed by the Psychological Corporation especially for the purpose of this study. The form was developed to obtain information about the child and the parents,
including education, occupation and race. A copy of this questionnaire is included in Appendix A.

Examiners

Three female and two male examiners participated in the study. All examiners were Caucasian and had completed at least two graduate courses in psychological assessment.

Procedure

Principals at the four schools were initially contacted requesting their schools' participation in this study. Parents were then sent a letter requesting their child's participation (see Appendix B), a consent form (see Appendix C), and the Background Information Questionnaire.

The children whose parents granted consent were randomly assigned to one of the examiners. Examiners individually administered the tests to the children, with 57 children receiving the WPPSI-R first and 47 children receiving the SB:FE first. The time between tests ranged from seven to 19 days, with a mean of 16 days. Children were tested at their school during the school day and were given appropriate breaks during testing.

Results

The question addressed in this study was the extent to which the WPPSI-R and the SB:FE are comparable measures of children's intelligence. Pearson product-moment
correlations between the SB:FE Composite score and the WPPSI-R Full Scale, Verbal, and Performance IQs were .46, .41, and .43, respectively. Correlations between the WPPSI-R Full Scale IQ and the SB:FE Area scores of Verbal Reasoning, Abstract Reasoning, Quantitative Reasoning, and Short-Term Memory were .44, .46, .39, and .42, respectively. The WPPSI-R Verbal IQ correlated most strongly with the SB:FE Verbal Reasoning Area at .55, and least strongly with the Short-Term Memory Area at .38. The Performance IQ and Area score correlations ranged from .44 for Quantitative Reasoning to .36 for Short-Term Memory. Table 1 demonstrates these relationships in further detail.

The mean Verbal, Performance, and Full Scale IQs on the WPPSI-R were 106.7 (SD = 12.71), 102.16 (SD = 11.19), and 104.18 (SD = 12.63), respectively. Mean scores on the SB:FE were 108.42 (SD = 12.22) for Verbal Reasoning, 107.04 (SD = 11.93) for Abstract/Visual Reasoning, 109.27 (SD = 12.83) for Quantitative Reasoning, 103.98 (SD = 11.69) for Short-Term Memory, and 106.84 (SD = 12.63) for the Composite score (see Table 2).
**Table 1**


<table>
<thead>
<tr>
<th></th>
<th>Verbal Reasoning</th>
<th>Abstract Reasoning</th>
<th>Quantitative Reasoning</th>
<th>Short-Term Memory</th>
<th>Composite Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object Assembly</td>
<td>.41</td>
<td>.42</td>
<td>.36</td>
<td>.39</td>
<td>.38</td>
</tr>
<tr>
<td>Geometric Design</td>
<td>.37</td>
<td>.46</td>
<td>.42</td>
<td>.38</td>
<td>.41</td>
</tr>
<tr>
<td>Mazes</td>
<td>.39</td>
<td>.39</td>
<td>.40</td>
<td>.36</td>
<td>.38</td>
</tr>
<tr>
<td>Picture Completion</td>
<td>.47</td>
<td>.39</td>
<td>.35</td>
<td>.37</td>
<td>.39</td>
</tr>
<tr>
<td>Animal Pegs</td>
<td>.36</td>
<td>.40</td>
<td>.36</td>
<td>.32</td>
<td>.35</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>.40</td>
<td>.43</td>
<td>.44</td>
<td>.36</td>
<td>.41</td>
</tr>
<tr>
<td>Information</td>
<td>.46</td>
<td>.37</td>
<td>.37</td>
<td>.42</td>
<td>.42</td>
</tr>
<tr>
<td>Comprehension</td>
<td>.51</td>
<td>.43</td>
<td>.43</td>
<td>.27</td>
<td>.43</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>.49</td>
<td>.46</td>
<td>.46</td>
<td>.43</td>
<td>.44</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.58</td>
<td>.56</td>
<td>.41</td>
<td>.38</td>
<td>.50</td>
</tr>
<tr>
<td>Similarities</td>
<td>.57</td>
<td>.51</td>
<td>.44</td>
<td>.40</td>
<td>.47</td>
</tr>
<tr>
<td>Sentences</td>
<td>.56</td>
<td>.53</td>
<td>.36</td>
<td>.37</td>
<td>.46</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>.55</td>
<td>.44</td>
<td>.39</td>
<td>.38</td>
<td>.43</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>.44</td>
<td>.46</td>
<td>.39</td>
<td>.42</td>
<td>.46</td>
</tr>
</tbody>
</table>
Table 2


<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WPPSI-R</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>106.7</td>
<td>12.71</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>102.16</td>
<td>11.19</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>104.18</td>
<td>12.63</td>
</tr>
<tr>
<td><strong>SB:FE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Reasoning</td>
<td>108.42</td>
<td>12.22</td>
</tr>
<tr>
<td>Abstract/Visual Reasoning</td>
<td>107.04</td>
<td>11.93</td>
</tr>
<tr>
<td>Quantitative Reasoning</td>
<td>109.27</td>
<td>12.83</td>
</tr>
<tr>
<td>Short-Term Memory</td>
<td>103.98</td>
<td>11.69</td>
</tr>
<tr>
<td>Composite Score</td>
<td>106.84</td>
<td>12.63</td>
</tr>
</tbody>
</table>

Discussion

The present data suggest that the WPPSI-R and the SB:FE are not comparable measures of intelligence in children. Correlations between the overall IQ's and subscales are consistently moderate. A 2.66 point difference was found between the mean Full Scale IQ and the mean Composite score. Moderate correlations
indicate the linear relationship between the two tests is not strong; the two tests are not interchangeable.

The weak correlations between scores on the two tests may suggest that they are measures of different, equally important aspects of intelligence. This multi-factoral model of intelligence proposes that intelligence is not unidimensional but is composed of many valid factors which may be measured by different tests. The correlations found in the present study, therefore, may not reflect a lack of validity but a lack of overlap in the constructs measured by the two tests.

The lack of support for the four area scores of the SB:FE may have affected the correlations in this study. Comparing WPPSI-R IQ’s to these scales may make interpretation difficult. The present investigator chose to use these scores because they may be used by some examiners and their psychometric properties should be examined. Further investigation comparing WPPSI-R scores to the SB:FE factor scores as suggested by Sattler (1988) would be useful in interpretation of the two measures.

The weak validity coefficients found in this study may also reflect in part the differences in administration procedures in the two tests. While the WPPSI-R relies more heavily on strict prompts, neutral examiner attitude, and controlled reinforcement, the SB:FE allows for greater
variation in these areas in order to obtain the child's best performance.

Obtaining a child's cooperation on tests often requires encouragement and occasional breaks, as intrinsic motivation may be lacking. Administrations in the current study involved breaks as needed by examinees, and therefore WPPSI-R scores may have been inflated by lack of adherence to standardization procedures. This may be operative in clinical administrations of the test as well and it should be considered when interpreting WPPSI-R scores.

Although the present study may not be used to determine the absolute utility of the WPPSI-R, it affords an examination of some factors that may affect clinical use of the test. The long administration time may be problematic and the new, colorful artwork may be distracting rather than simply engaging for some children. Additionally, the stringent standardization procedures may result in a greater likelihood of nonstandard administrations, resulting in questionnable scores.

The moderate correlations found in this study along with the possible problems in the WPPSI-R as discussed above bring into question whether or not the new test is an improvement over the original version. The present study does not provide enough information to make a definitive
statement about the validity of the revision or the advisability of its use over that of the WPPSI. Until the aspects of intelligence measured by the WPPSI-R are better understood, clinicians should carefully interpret results. The present study should be considered one step in the process of determining the usefulness of the WPPSI-R.
APPENDIX A

BACKGROUND INFORMATION QUESTIONNAIRE
APPENDIX B

LETTER TO PARENTS OF PARTICIPANTS
APPENDIX C

CONSENT FORM
REFERENCES


