1992

A Comparison of the Performance of African American Preschool Children on the WPPSI and WPPSI-R

Cynthia A. Vandervelde
University of Northern Iowa

A COMPARISON OF THE PERFORMANCE OF AFRICAN AMERICAN PRESCHOOL CHILDREN ON THE WPPSI AND THE WPPSI-R

An Abstract of a Thesis

Submitted

In Partial Fulfillment

of the Requirements for the Degree

Specialist in Education

Cynthia A. Vandervelde

University of Northern Iowa

July 1992

ABSTRACT

The Wechsler Preschool and Primary Scale of Intelligence (WPPSI)(Wechsler, 1967) has long been one of the most widely used instruments for the assessment of preschool children's intelligence. The recently published Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R)(Wechsler, 1989) was designed to replace the WPPSI as the instrument of choice for the purposes of making classification, placement, and program evaluation decisions. Thus, data concerning the comparability of the two instruments is of particular importance to educators. This study investigated the relationship between the WPPSI and the WPPSI-R among a sample of predominantly lower socioeconomic status African American children in order to afford some estimate of the relative similarities or differences that might exist between the two instruments. Twenty-four African American children ages four and five were administered the WPPSI and the WPPSI-R in counterbalanced order. Group mean scores on the individual subtests and on the Verbal, Performance, and Full Scales of the two instruments were compared. Results indicated that scores on the WPPSI were consistently higher than were those on the WPPSI-R. Differences between the two tests were 8.3, 13.6, and 11.7 in favor of the WPPSI for the Verbal, Performance, and Full Scale IQs,

respectively. Data from this study support the view that population gains in IQ scores are occurring at the lower socioeconomic strata. The practical implications of changes in IQ scores due to the re-norming of the WPPSI-R are discussed.

A COMPARISON OF THE PERFORMANCE OF AFRICAN AMERICAN PRESCHOOL CHILDREN ON THE WPPSI AND THE WPPSI-R

A Thesis

Submitted

In Partial Fulfillment

of the Requirements for the Degree

Specialist in Education

Cynthia A. Vandervelde

University of Northern Iowa

July 1992

This study by: Cynthia A. Vandervelde

Entitled: A Comparison of the Performance of African American Preschool Children on the WPPSI and the WPPSI-R

has been approved as meeting the thesis requirement for the Degree of Specialist in Education.

7/15/92
Date — Dr. Ralph S. Scott, Chair, Thesis Committee

7/15/92
Date — Dr. Melissa L. Heston, Thesis Committee Member

7/15/92
Date — Dr. Donna B. Raschke, Thesis Committee Member

7-31-92
Date — Dr. John W. Somervill, Dean, Graduate College

ACKNOWLEDGEMENTS

I would like to thank my committee members, Dr. Ralph Scott, Dr. Donna Raschke, and Dr. Melissa Heston for their support, knowledge, and patience.

The contributions of Mike Knapp and his staff at Grin and Grow Day Care were instrumental to the success of this study. I would also like to thank Kelly Kincaid and her staff at Lily Ferguson, and Mary Powell and her staff at St. Mary's. My gratitude also goes out to the many delightful children I tested, who made the data collection aspect of this project an incredibly fun and rewarding process.

Worthy of special thanks is my friend Matthew B. Green, without whose continual guidance and support I would not have completed this thesis. Tim Harmon also deserves thanks for contributing his knowledge of statistics and computer programming.

I would also like to thank Patricia Heady for sharing her word processing knowledge and skills with me.

Finally, I want to thank my son, Julian Zebedee Vandervelde, who has given up untold hours of "mother time" in order to allow me this personal accomplishment, and who has been my mainstay throughout.

TABLE OF CONTENTS

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

When a major psychometric instrument is revised questions inevitably arise concerning the effects of the revision on the meaning of the scores. When the instrument's use is as widespread as that of the Wechsler preschool scales, the comparability of the original and revised instruments is of considerable importance to educators. This is especially true in light of recent legislation, most notably P.L. 102-119, The Individuals with Disabilities Education Act Amendments of 1991, which mandates the early identification and assessment of 3 to 5 year old children who may later experience learning problems. Intelligence tests are frequently an important part of the multifactored process whereby students are identified as exceptional and become eligible for preventive, remedial, or compensatory educational programming. Instruments such as the Wechsler Preschool and Primary Scale of Intelligence (WPPSI)(Wechsler, 1967) and the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R)(Wechsler, 1989) are used by psychologists to describe patterns of behavior, identify cognitive strengths and weaknesses, and develop effective pedagogical strategies for addressing learning

difficulties. Therefore, it is important that the relationship between the WPPSI and the WPPSI-R be examined empirically.

Questions concerning the comparability of the WPPSI and the WPPSI-R also pertain to a concern surrounding the use of nonbiased assessment techniques. Throughout the 1970s and 1980s a central issue in the literature and in litigation (e.g., Larry P. v. Riles, 1984; PASE v. Hannon, 1980) has been the fairness of intelligence tests for minority students. The disproportionately high rate of African American children in classes for the mildly handicapped has fueled this debate (Hilliard, 1979; Reschly, 1981). The fairness of tests to minority youth is a complex problem to which psychologists have devoted considerable study. Several technical indices of racial bias show no appreciable indication of such in the WPPSI when used with African American children (Clarizio, 1978; Kaufman, 1973; Quereshi & McIntire, 1984). However, a considerable body of research is accruing which indicates that more recent forms of IQ tests yield lower IQ scores (Flynn, 1984). From a practical perspective, such findings may mean that children must perform at a higher level on newly standardized tests in order to achieve scores equivalent to those obtained on an earlier form. If this is indeed the case, then the revised

WPPSI-R may significantly affect the incidence of African American children classified for special education programs.

From a theoretical standpoint, the data presented by Flynn (1984, 1987) and others who are currently studying the change in ability levels over generations challenges the commonly accepted notion of fixed intelligence. The methodology used in the studies cited by Flynn involves the administration of two versions of an IQ test, normed at different times, to the same subjects. Support for an average three point per decade increase in American IQ scores since 1932 has been reported (Flynn, 1984; Lynn & Hampson, 1986; Teasdale & Owen, 1989). The group data presented by these investigators demonstrate the malleability of IQ. Research such as this, conducted on different cultures over time (or retrospectively) may isolate specific environmental factors associated with gains in intelligence. Such research could lead to the development of successful interventions to reduce group IQ differences between races and across social classes.

Several studies have recently appeared in the literature examining the concurrent validity of the original and revised Wechsler preschool scales. The manual for the WPPSI-R (Wechsler, 1989) includes one study comparing the scores from the WPPSI-R and the WPPSI. In

addition, Kaplan, Fox, and Paxton (1991) and Milrod and Rescorla (1991) have reported studies comparing the two versions of the test with Caucasian, high socioeconomic status youth. Each of these studies found differences between the two instruments on the Verbal, Performance, and Full Scale IQs, with the WPPSI means being significantly higher. Further research is needed to demonstrate the comparability of these tests (and their norms) with lower socioeconomic status and minority populations.

Statement of the Problem

As school psychologists become more involved in the assessment of preschool children and the development of educational programs for these young people, it becomes increasingly important that the instruments they use be of the highest quality, and valid for all populations to which they are administered. The validity of standardized tests of mental ability is of utmost importance since the accurate estimate of a child's intellectual functioning is often crucial to his or her educational placement and subsequent development.

The present study is designed to determine the comparability of the WPPSI and the WPPSI-R for a group of children in the age overlap of 4 to 5 years, from a

predominantly lower socioeconomic status African American population. The following questions are asked:

1. Are there significant differences in Verbal IQ, Performance IQ, and Full Scale IQ scores on the WPPSI and the WPPSI-R scales?

2. What are the interrelationships between the IQ scores derived from the two scales?

3. Do the data from this study support research indicating large gains in the IQ scores of Americans over the last 50 years?

Research Hypothesis

There will be significant differences in the Verbal, Performance, and Full Scale IQ scores and in the subtest scaled scores between the WPPSI and the WPPSI-R with African American children.

Null Hypothesis

There will be no difference in the Verbal, Performance, and Full Scale IQ scores and in the subtest scaled scores between the WPPSI and the WPPSI-R with African American children.

Definition of Terms

Intelligence

An individual's overall capacity to understand and cope with the world around him or her. David Wechsler (1950) viewed intelligence as multifaceted and

multidetermined, including such elements as abstract reasoning, verbal, spatial, and numerical factors, and drive and incentive.

African American

Those individuals, born in the United States of America, of Sub-Saharan African descent. These individuals have variously been referred to in the literature as negro, black, and Afro-American.

Validity

The concern with the extent to which an instrument measures what it purports to measure. It is always specific to the particular purpose for which the instrument is used. There are several ways to assess the validity of an instrument (Salvia & Ysseldyke, 1988).

1. Content validity shows the extent to which a sample of items on a test is representative of some defined domain of content. Content validity is established by examining three factors: item appropriateness, item completeness, and the way in which items assess the content.

2. Criterion validity refers to the extent to which a person's score on a criterion measure can be estimated from that person's test score. Concurrent criterion validity refers to how accurately a person's test score can be used to estimate the current criterion score.

Predictive criterion validity refers to how accurately a person's test score can be used to estimate what the criterion score will be at a later time.

3. Construct validity is the extent to which a test measures a theoretical trait or characteristic (e.g., intelligence). In order to validate a test of a construct, scientists rely on indirect evidence and inference.

Bias in Testing

The American Psychological Association (Walker, 1991) defines test bias as "any significant differential performance on tests by different populations (e.g., males v. females) as a result of test characteristics which are irrelevant to the variable being measured" (p. 220). The inference of bias is made only after a great deal of evidence regarding the tests' content, construct, and criterion-related validity has been weighed.

The Wechsler Scales

Clinical intelligence test batteries for preschool children, school age children, and adults, assembled by David Wechsler since the 1930s. These scales have been modified and expanded through various revisions over the years. The abbreviation WPPSI is used throughout to indicate the Wechsler Preschool and Primary Scale of Intelligence, while the WPPSI-R refers to the 1989

revision of that instrument. WISC signifies the Wechsler Intelligence Scale for Children, whereas WISC-R indicates the 1969 revision, and WISC III signifies the 1990 revision.

Lower Socioeconomic Status

Those families who meet the State of Iowa guidelines for subsidized funding of child care. The Iowa Department of Human Services determines eligibility for subsidized child care via a formula which includes family size and gross monthly income. For example, a four member family with an income of less than $1,187.00 dollars per month is eligible for subsidized child care.

CHAPTER II

REVIEW OF THE LITERATURE

This review of the literature will be organized as follows:

I. Increases in IQ over generations

II. Comparisons of original and revised versions of the Wechsler scales.

III. The Wechsler Preschool and Primary Scale of Intelligence (WPPSI)

- A. Description and standardization
- B. Validity
  1. construct
  2. concurrent
  3. predictive

IV. The Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R)

- A. Description and standardization
- B. Validity
  1. construct
  2. concurrent

V. Validity of the WPPSI and the WPPSI-R with African American children

A. Content validity

B. Construct validity

C. Criterion-related validity

Increases in IQ Over Generations

A substantial body of research is accumulating that has direct bearing on the issue of IQ malleability. Flynn (1984), Lynn and Hampson (1986), and Teasdale and Owen (1989) have compared IQ scores yielded by the older versus the revised editions of intelligence tests in order to measure IQ trends from one generation to the next. The combined results of their analyses indicate that in a number of economically developed nations the intelligence of the population has increased by approximately one standard deviation over the last half century. These increases have been reported in the United States by Flynn (1984) who has compared the IQ scores yielded by the original and revised batteries of the Stanford-Binet and the Wechsler scales of intelligence. By analyzing the data of 73 studies including more than 7,500 subjects between the ages of 2 and 48 years, Flynn found that when the same subjects took two Stanford-Binet or Wechsler tests, they almost invariably got lower scores on whichever test had been standardized at a later date.

Flynn reported that Americans have gained almost 15 IQ points between 1932 and 1978. This seeming paradox occurs because more recent standardization samples do better on the same test questions than did the previous standardization samples. In practical terms, this means that an individual obtains lower scores on a restandardized test than on a test normed earlier, and on a less able, population. Similar group increases in IQ have been documented for Japan, Britain, Australia, New Zealand, Holland, and other Western European countries (Flynn, 1987; Lynn, 1990; Lynn & Hampson, 1986), although variations in the magnitude of gain vary from country to country. Also noted in these studies is that the size of IQ gain seems to be related to the type of IQ test employed: tests of visuo-spatial abilities, or those considered more "culture free," tend to show a greater increase than do those of verbal-educational abilities (Flynn, 1987; Lynn, 1990; Lynn & Hampson, 1986).

Researchers have attempted to explain these population increases in IQ in several ways. The reach of mass media and television in particular, improved and better-informed parenting techniques, increases in the quality of education and the years of formal schooling achieved in the general population, and improved nutrition have all been hypothesized as possible explanatory factors (Flynn,

1987; Lynn & Hampson, 1986). Flynn (1987) states that based on reproductive potentials, it is virtually impossible to attribute the observed gains to genetic factors. He also calls into question the construct validity of intelligence tests, arguing that the increases in intelligence cannot be genuine, and therefore intelligence tests do not measure intelligence, but rather "abstract problem solving ability" (p. 189). Concurrent declines in Scholastic Aptitude test scores in the United States also call into question the reasonableness of assuming a general population increase in intelligence. A recent study by Rosenbach and Rusch (1991) correlated student performance on a 1930's version of the Pinter General Ability Tests: Verbal Series with several more current measures of intelligence and achievement. The mean subject performance on the tests compared did not differ significantly, leading the authors to conclude that their data fail to show generational increases in intelligence. Nonetheless, each revision of the Wechsler scales has been empirically shown to produce lower mean IQ scores than its predecessor.

Comparisons of the Original and Revised Versions of the Wechsler Scales

Brooks (1977) compared scores on the Wechsler Intelligence Scale for Children (WISC) (Wechsler, 1949)

and the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) and found group mean scores on the Verbal, Performance, and Full Scales to be between 3 and 11 points lower on the revised instrument. Similarly, Catron and Catron (1977) compared the performance of educable mentally retarded students on the WISC and the WISC-R and noted that the students performed approximately 5 to 6 points lower on the WISC-R. The Wechsler Adult Scale of Intelligence (WAIS)(Wechsler, 1955) was compared to the revised edition published in 1983 and subjects were again found to perform significantly lower on the more recently standardized test. The authors concluded that given the large common item pool, stable differences between forms, and the high correlations found between the scales and the subtests, the normative samples must necessarily be very different (Lippold & Claiborn, 1983).

The most recently restandardized version of the Wechsler scales, the WISC III, was compared to its predecessor the WISC-R in a study reported in the manual (Wechsler, 1991). Not surprisingly, a comparison of the mean Full Scale IQ scores showed that the WISC III Full Scale IQ was approximately 5 points less than that of the WISC-R. The mean WISC III Verbal and Performance IQs were approximately 2 and 7 points less than the corresponding WISC-R scores, respectively.

Finally, in those studies where the mean scores of the WPPSI and WPPSI-R were compared, the mean WPPSI-R scores were consistently lower than were those of the older version of the instrument (Griggs, 1991; Milrod & Rescorla, 1991; Wechsler, 1989). Thus, there exists a wealth of data supporting generational scoring differences on the Wechsler scales that is not adequately explained by minor changes in content and administration in the revised editions of the instruments.

The WPPSI, Description, and Standardization

David Wechsler originally developed the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) (Wechsler, 1967) to meet the need for a measure of intellectual ability that extended below the age range of the Wechsler Intelligence Scale for Children (WISC) (Wechsler, 1949; Wasik & Wasik, 1972). Oldridge and Allison (1968) described the WPPSI as a carefully developed and well standardized measure of general intelligence.

The WPPSI consists of a battery of subtests, each of which when treated separately may be considered as measuring a separate cognitive strength or weakness (Carlson & Reynolds, 1981), and when combined into a composite score measure overall intellectual capacity. The WPPSI taps many developmental or maturational factors

particularly important to school success in the lower grades. Of the 11 subtests, 10 are used to calculate the IQ. Subtests are grouped into a verbal and a performance section, from which Verbal, Performance, and Full Scale IQs are obtained. The verbal and performance subtests are presented in alternating order to encourage cooperation and interest in young children. In scoring the WPPSI, subtest raw scores are converted to scaled scores, which are then converted to normalized deviation IQs with a mean of 100 and a standard deviation of 15.

The WPPSI was standardized on a nationwide sample of 1,200 children stratified against the 1960 census data with respect to geographical region, urban-rural residence, father's occupation level, and proportion of whites and nonwhites (Wechsler, 1967). IQ scores obtained are a comparison of the examinees' scores with the scores earned by a representative sample of her or his own age group.

Bracken (1987) evaluated 10 commonly used preschool assessment instruments,including the WPPSI, after establishing technical adequacy criteria in the areas of total test and subtest internal consistency, test-retest reliability, floor adequacy, steepness of subtest item gradients, and the presence of validity information. The WPPSI was the only instrument of those evaluated which

consistently met the criteria for adequacy, although the subtest floor was considered inadequate through age 4 years 6 months. Bracken concluded that the WPPSI provides reliable and valid estimates of preschool children's abilities.

Construct Validity of the WPPSI

The most frequently used method for measuring construct validity has been factor analysis. The factor analytic approach is a statistical technique used to identify clusters of interrelated items on a test. These clusters of items are then examined for common features and given a name based on the construct reflected. Factor analysis helps identify underlying, not directly observable, constructs. A number of factor analytic studies have been conducted on the WPPSI (Carlson & Reynolds, 1981). Two factors have consistently been found, corresponding essentially to Wechsler's a priori grouping of the subtests into a Verbal and Performance Scale. The separate interpretation of verbal and performance abilities is consistent with Wechsler's theory and with current practice of interpreting the WPPSI (Gyurke, Stone, & Beyer, 1990).

Concurrent Validity of the WPPSI

By any measure, the bulk of studies investigating the validity of the WPPSI are those in which the instrument is

correlated with some other measure of intelligence. Correlations between the WPPSI and other measures of mental ability are substantial and significant in almost every case. The WPPSI can be said to have good concurrent validity with other intellectual measures. Several studies have compared the WPPSI and the WISC and found them to correlate significantly (Austin, 1970; Oakland, King, White, & Eckman, 1971). In a well designed and executed study, Quereshi and McIntire (1984) compared seventy-two 5 and 6-year-old children on the WPPSI, WISC, and the WISC-R using four separate criteria for parallelism: equality of means, variances, reliabilities, and validities. Their results indicated that Verbal IQs on all three instruments were comparable, but the Performance and Full Scale IQs were not. Among the subtests common to all three instruments, only Comprehension, Arithmetic, Picture Completion and Mazes satisfied all of the criteria of parallelism set forth in the study. The authors hypothesized that the observed discrepancy in performance IQ across the three scales was due to differences in the content among the performance subtests.

Kutsick and Wynn (1988) compared another test of intelligence, the Kaufman Assessment Battery for Children (K-ABC) (Kaufman and Kaufman, 1983), and the WPPSI.

Strong positive correlations (r = .66 to .79) were found between the K-ABC Total Achievement Scale and the WPPSI Verbal, Performance, and Full Scale IQs.

Oakland et al. (1971) investigated the concurrent validity of the WPPSI, WISC, and Stanford Binet with a sample of lower socioeconomic status African American children. The WPPSI Full Scale IQ correlated significantly with both the WISC Full Scale IQ (r = .65) and the Stanford Binet (r = .74). However, the authors noted that the subjects performed significantly higher on the WISC and the Binet than on the more recently developed WPPSI. Results such as these have occurred repeatedly in the literature. While many intelligence tests correlate at a highly significant level, thus indicating their concurrent validity, in almost every case students score lower on more recently standardized instruments.

Another study (Kutsick, Vance, Scwarting, & West, 1988) compared the performance of preschool children identified as "at risk" on three measures of intelligence: the WPPSI, the Peabody Picture Vocabulary Test-Revised (PPVT-R), and the Expressive One Word Picture Vocabulary Test (EOWPVT). Significant correlations between these three instruments were obtained: The WPPSI Full Scale IQ correlated with the PPVT-R at the .85 level, and with the EOWPVT at the .78 level. The authors concluded that the

three instruments have adequate concurrent validity for at risk preschool children.

Predictive Validity of the WPPSI

Anastasi (1968) points out that the basic difference between concurrent and predictive validity is the objective of the testing. Predictive validity is concerned with the prediction of future outcomes. A number of studies have been conducted in which the WPPSI has been used to predict future intellectual and academic performance. The bulk of these studies indicate that the WPPSI adequately predicts achievement test scores and academic performance. Crockett, Rardin, and Pasewark (1976) administered the Metropolitan Achievement Test (MAT) to 35 children who had been tested with the WPPSI 4 years before. Correlations between the WPPSI Verbal IQ and the subject areas of the MAT were not significant, while the Performance IQ correlated only moderately with the mathematical components of the MAT (r = .42 to .52). Conversely, White and Jacobs (1979) found significant moderate correlations between WPPSI IQs and later reading achievement (r = .51 to .58) for a sample of middle class nursery school children. Several other studies have reported a positive relationship between WPPSI IQs of kindergarten children and later reading achievement (Hagin, Silver, & Corwin, 1971; Lieblick & Shinar, 1975).

In a 12 year longitudinal study, Lowe, Anderson, Williams, and Currie (1987) administered the WPPSI to 169 culturally deprived African American children prior to first grade. These scores were then correlated with the Metropolitan Readiness test (MRT), the Iowa Test of Basic Skills (ITBS), the Children's Achievement Test (CAT), and the Iowa Tests of Educational Development (ITED) when the original sample of children reached 1st, 5th, 8th, and 11th grade. The subject's WPPSI scores were also correlated with their grade point averages at each grade level. A substantial portion of the correlations found in the study were highly significant ($p<.001$). For example, the correlation between overall grade point averages and WPPSI Full Scale IQs for grades 1 through 11 ranged from .45 to .64. The WPPSI correlated with the MRT and ITBS composite scores at the .45 level, and with the ITED scores at the .30 level. These findings support the view that for lower socioeconomic African American youth, the WPPSI has long range validity in predicting both school performance and achievement test scores. In summary, numerous studies have found that performance on the WPPSI correlates significantly with other tests of mental ability, and adequately predicts both academic performance and achievement test scores. The WPPSI demonstrates both concurrent and predictive validity.

The WPPSI-R, Description and Standardization

The Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R)(Wechsler, 1989) is a thorough revision of the WPPSI, retaining 61.7% of the items from the original instrument (Kaufman, 1990a). Principle changes in the revised edition include the addition of a new subtest (Object Assembly), an expanded age range (3 to 7 years), new colorful multicultural materials, new item types for very young children, and some changes in administration procedurers. Despite these changes, the WPPSI-R is similar to the WPPSI in terms of organization, focus, and application.

The standardization sample for the WPPSI-R was composed of 1,700 children, and like its predecessor was stratified based on census proportions on the variables of age, gender, geographic region, ethnicity (whites, blacks, Latinos, and others), parent's education, and parent's occupation (Wechsler, 1989).

Construct Validity of the WPPSI-R

According to the WPPSI-R manual, much of the literature regarding the validity of the WPPSI is also relevant to the WPPSI-R. Like the WPPSI, studies examining the construct validity of the WPPSI-R have employed a factor analytic approach. Factor analysis has supported the interpretation of a separate verbal and

performance factor (coinciding generally with the verbal and performance scales) as well as an underlying general component (Gutkin & Reynolds, 1990; Wechsler, 1989).

Concurrent Validity of the WPPSI-R

The WPPSI-R manual (Wechsler, 1989) cites a number of concurrent validity studies, including comparisons of the WPPSI-R to the WPPSI, WISC-R, Stanford-Binet Intelligence Scale-Fourth Edition, McCarthy Scales of Children's Abilities, and Kaufman Assessment Battery for children (K-ABC). In addition, the newly published WISC III has been compared to the WPPSI-R (Wechsler, 1991).

In a sample of 50 children, the WPPSI-R yielded lower Full Scale group mean IQ scores than the WISC by approximately 8 points. The correlation coefficient of the two Full Scale IQ scores was .85. When compared to the Stanford-Binet Intelligence Scale-Fourth Edition, the WPPSI-R full scale IQ was about 2 points lower than the mean composite score on the Stanford-Binet. The correlation coefficient was .74 (Wechsler, 1989).

The WPPSI-R yielded scores that were similar to those on the McCarthy Scales of Children's Abilities, although again they were approximately 2 points lower. The two instruments correlated highly (r = .81) (Wechsler, 1989).

The Kaufman Assessment Battery for Children and the WPPSI-R correlated only moderately (r = .49), and the mean

WPPSI-R Full Scale IQ was about 6 points lower than the mean K-ABC mental composite.

In a concurrent validity study of the WPPSI-R and the WPPSI (Wechsler, 1989), 144 children between the ages of 48 and 79 months were administered the two tests in counterbalanced order. The correlations for the Performance, Verbal, and Full Scale IQs were reported as .82, .85, and .87 respectively, thus indicating that the two instruments measure the same constructs. However, comparison of the group mean IQ scores showed that the WPPSI full scale IQ was approximately eight points higher than that of the WPPSI-R. The WPPSI Performance and Verbal IQs showed a similar trend (9 and 5 points higher, respectively, for the WPPSI). Griggs (1991) reported similar results in a comparative study analyzing the WPPSI and WPPSI-R scores of 35 students. Consistent with the findings reported in the previous study, the group mean Verbal, Performance, and Full Scale IQ scores of the revised test were lower than those of the original by 3.1, 11.7, and 7.8 points, respectively. Correlations between each of the scales were found to be high, although somewhat lower than those cited in the manual: Verbal Scale = .60, Performance Scale = .62, Full Scale = .71.

Milrod and Rescorla (1991) compared the performance of 80 high socioeconomic status students on the WPPSI and

the WPPSI-R and found that the two instruments produce comparable results. The WPPSI/WPPSI-R Performance IQ, Verbal IQ, and Full Scale IQ comparisons all produced correlation coefficients equal to or greater than .70. WPPSI-R Full Scale, Performance, and Verbal IQs were 6.0, 9.2, and 3.5 points lower, respectively, than were their WPPSI counterparts.

The WISC III, designed for the assessment of children ages 6 through 16, has less item overlap than do the WPPSI and the WPPSI-R. Nonetheless, the two tests correlate significantly on all three scales (Verbal Scale = .85, Performance Scale = .73, Full Scale = .85). Somewhat surprisingly, the group mean IQ scores on the WPPSI-R ranged from 2 to 6 points lower than those on the more recently normed WISC III (Wechsler, 1991). The authors attribute this difference to the fact that the WISC-III has a higher ceiling for above average 6 and 7 year olds. They suggest that for this reason the WISC-III is the better instrument for children in this age range.

Validity of the WPPSI and the WPPSI-R with African American Children

The issue of the validity of the WPPSI-R also pertains to the determination of whether this instrument is fair or biased when used with African American children. The issue of bias in intelligence tests and the

inequitable social outcomes of such testing for culturally diverse children has generated heated controversy over the last two decades (Reynolds & Hartlage, 1979; Reynolds, Wright, & Dappen, 1981). Validity criteria are generally seen as the most legitimate criteria for judging test bias. The American Psychological Association's committee on the use of education and psychological tests with minority children has incorporated all three forms of validity (content, construct, and criterion-related) as important variables in the issue of bias (Cleary, Humphreys, Kendrick, & Wesman, 1975).

Content Validity with African Americans

Evaluation of content bias focuses on whether a particular test item is unfair to some groups of the population. Specific test items that systematically favor one group over another are considered biased (Gutkin & Reynolds, 1990). In the development of the WPPSI-R, two methods were used to analyze potential content bias with respect to race/ethnicity (Wechsler, 1989). One method can be described as an analysis based on the "face validity," or subjective validity, of the test items. A panel of expert reviewers, including psychologists familiar with ethnic bias studies and members of the Psychological Corporation, identified items they felt were potentially biased against race, ethnicity, or sex. These

items were then reconstructed to minimize the perceived sources of bias, or were eliminated from the test (Wechsler, 1989).

A second, empirical method utilized to eliminate item bias on the WPPSI-R was conducted using a sample including 400 minority children. Items whose difficulties varied by subgroup (Caucasian, African-American, and Hispanic) were identified. Next, a partial correlation technique was applied to detect between-group differences in item performance after controlling for overall subtest performance. Items identified as biased through this analysis were discarded from the final version of the WPPSI-R. This method of test development, using both reviewer evaluation and empirical analysis in order to reduce content bias in the instrument, has been promoted by Berk (1982) and is accepted by many professionals as the most equitable form of test development (Gutkin & Reynolds, 1990).

Other than the above, studies examining the cross-racial validity of the WPPSI-R have yet to appear in the literature. However, the WPPSI has been subjected to a great deal of empirical analysis regarding bias. Additionally, studies conducted on the WISC-R (an upward extension of the WPPSI) are relevant to the present analysis of content bias due to the similarity of the

three Wechsler scales in terms of item, subtest, administration, and age overlap.

Meile (1979) investigated the possible content bias of WPPSI with a sample of 274 children (111 African American and 163 Caucasian). Race by item interaction was analyzed to determine if rank order of item difficulty was significantly different in the two populations. The results of his study indicated that item difficulty was essentially the same across race.

Sandoval (1979) examined cross-racial correlations of item difficulty on the WISC-R using a sample of 1,050 children from three ethnic groups (Caucasian, African American, and Hispanic). Rank order correlations between item difficulties for the three groups were all quite high, clustering around .98, indicating that few items are relatively more difficult for one group or another. Sandoval concluded that, "the internal criteria of test bias against minority children has not been found for the WISC-R"(p. 926). In another study, Reynolds (1980) examined each of 12 WISC-R subtests for cultural bias against African Americans. The WISC-R standardization sample was used to match 270 African American and Caucasian children on the basis of Full Scale IQ. Group differences in performance on each of the subtests were then examined to determine which of the subtests were

disproportionately difficult. Reynolds found significant differences in subtest performance between the two groups on several subtests: African Americans exceeded Caucasians on the Digit Span and Coding subtests, and Caucasians exceeded African Americans on the Comprehension, Object Assembly, and Mazes subtests. While these results can be interpreted to indicate bias in several of the WISC-R subtests, the amount of variance associated with ethnic group membership was less than 5% in each case, and therfore of little practical significance. Thus, a number of studies have accumulated presenting a relatively clear picture when taken as a whole: Differential item bias across race has not been supported by research on the Wechsler scales.

Construct Validity with African Americans

A second form of test validity related to bias, construct validity, assumes that scores from a test given to different cultural groups measure the same abilities for the various groups. Indeed, if the test is not measuring the same underlying abilities or if the commonly used scores from the test represent varying abilities depending on group membership, then use of the test with culturally different persons is inappropriate and unfair (Reynolds, 1980). The most frequently used approach to measuring construct validity has been factor analysis.

For example, Kaufman and Hollenbeck (1977) used the WPPSI standardization sample to compare the structure of the WPPSI for groups of African American and Caucasian children. The two groups were stratified in accordance with the 1960 census data on the variables of age, sex, geographic region, father's occupation, and urban-rural residence. Results indicated that the WPPSI has virtually the same factor structure for African American and Caucasian children: verbal and performance factors emerged and correlated at .84 and .87 for the two groups, respectively. The authors concluded that the WPPSI is a fair instrument when judged on the criteria of assessing the same underlying abilities for both groups of children.

Using a large random sample, Reschly (1978) compared the factor structure of the WISC-R across four racially diverse groups: Caucasian, African American, Chicano, and Native American Papago. As in the Kaufman and Hollenbeck (1977) study, substantial congruency of factors across race were found when two factor (verbal and performance) solutions were compared. Coefficients of congruence across all combinations of racial groupings were .97 to .99, indicating equivalence across groups. A general intelligence factor was also found to be congruent across race, leading Reschly to conclude that the WISC-R Full Scale IQ as a measure of overall ability as well as the

verbal and performance distinctions, are equally appropriate for African Americans, Caucasians, Chicanos, and Native American Papagos.

Gutkin and Reynolds (1981) employed the WISC-R standardization sample to investigate the factorial similarity across race. Coefficients of congruence were .99 for a general "g" factor, a two factor, and a three factor solution across race. For African American and Caucasian children, the authors concluded, the WISC-R factor structure is extremely similar.

Overall, it can be seen that there is a considerable body of evidence indicating a lack of construct bias in the Wechsler scales, thus supporting uniform interpretation of Full Scale, Verbal, and Performance IQs independent of race.

Criterion-Related Validity with African Americans

Investigations of the differential predictive validity of IQ scores examine the regression equations for different groups in the prediction of grades, teacher ratings, and achievement test scores from IQ scores. Generally, studies investigating WISC-R IQs and school achievement test scores have found IQ to be a valid predictor of achievement regardless of ethnicity (Hartlage & Steele, 1977; Reschly & Reschly, 1979). However, the relationship of IQ to other measures of academic ability

has not been so straightforward. Reschly and Reschly (1979) have proposed the need to further investigate the predictive validity of the Wechsler IQs for African American children relative to grades. However, skepticism has been voiced regarding the appropriateness of using academic grades as criterion measures of validity due to the variability of grading practices among teachers and between school districts (Lowe et al., 1987).

Goldman and Hartig (1976) investigated the relationship among WISC IQs, grade point averages (GPA), and teacher ratings of academic competence. They reported that IQ was a poor predictor of GPA and teacher ratings for minority students, and only a moderate predictor for Caucasian children. The authors concluded that the WISC IQ is not differentially valid for Caucasian and minority children when classroom performance is used as the criterion measure.

Partenio and Taylor (1985) sent a 4-item rating scale concerned with classroom performance, motivation to learn, and learning potential to the teachers of 120 randomly chosen students (40 Caucasian, 40 African American, and 40 Hispanic) who had been administered either the WPPSI or the WISC-R during a restandardization project. Scores from the teacher rating scale were correlated with Full Scale, Performance, and Verbal IQs for each ethnic group.

The resulting data indicated that the relationship between IQ and teacher ratings was stronger for Caucasian than for minority students. These authors proposed two explanations for their findings: first, that the Wechsler scales are not valid predictors of scholastic performance because of differential biasing effects, and second, that teacher ratings are biased measures of scholastic competence. They concluded that it is impossible to determine which of these two explanations is true.

Data from the empirical investigation of test bias suggest several guidelines for the equitable assessment of all children: investigation of possible referral source bias, inspection of the test developer data regarding statistical analysis of bias, assessment using the most reliable measures available, and the use of multiple sources of data prior to making decisions concerning children (Gutkin & Reynolds, 1990).

Assessment procedures aside, clinicians and educators need to be aware of the well-documented finding that African Americans and Caucasians as a group do differ in their mean IQ scores. This difference amounts to approximately 10 IQ points when the groups are matched on socioeconomic status (Kaufman & Hollenbeck, 1977; Reynolds & Hartlage, 1979; Sattler, 1988). These differences have been variously attributed to genetic endowment and

environmental opportunity, with the controversy surrounding the relative importance of each being intense and heated. Psychologists and educators who use IQ tests generally reject genetic explanations of group differences. Nonetheless, overrepresentation of African American students in programs for the mildly retarded has been tied to racist views of intellectual potential (Hilliard, 1980). Angoff (1988) has argued that the wrong question has continually been asked by those trying to determine the relative influences of heredity and environment on IQ variability. He states:

> The real issue is whether intelligence can be changed, an issue that does not at all go hand in hand with the issue of heritability. Whatever the "true" heritability coefficient for intelligence is . . . , the essential point is that in the context of group differences and what these differences connote, its numerical value is irrelevant. What is relevant is whether these group differences can be changed, with what means, and with what effect. (p. 716)

Kaufman (1990b) states that a substantial body of data currently point to the concept of malleability of intelligence for whole cultures and confirm the fact that intelligence tends to be in flux. The average intelligence of Americans seems to be increasing at a steady, and measurable rate. Research conducted on different cultures over time may help isolate specific environmental factors associated with these gains, and could have far reaching implications for the interventions

targeted not only at the individual child, but also at the entire community.

CHAPTER III

METHODOLOGY

Subjects

The subjects were 24 predominantly lower socioeconomic status African American children selected from three inner city preschool centers in a midwestern metropolitan area with a population of 105,000. The majority of children enrolled in the three preschools receive Iowa Department of Human Services (DHS) subsidized funding of child care expenses: in the first preschool 99% of the children are DHS subsidized, in the second 93%, and in the third 66%. Nine of the subjects were enrolled in a federally funded program for at-risk preschool children. Eligibility criteria for the program included family income under 125% of the federal poverty guidelines, identified developmental delays, premature birth, parental illiteracy, or parental history of substance abuse or chronic mental illness. It should be noted that information regarding the economic status of individual subjects in the study was not available.

The subjects ranged in age from 48 through 71 months (M = 54) at the time of the initial testing; 12 were male and 12 were female. The original sample pool consisted of 30 children, however, due to absences or refusal to comply with test procedures, the results of 6 children (4 males

and 2 females) were not deemed appropriate for use in the study. Participation in the study was voluntary, and informed consent was obtained from the parents of each of the subjects (see Appendix A).

Instruments

WPPSI

The WPPSI is an intelligence test consisting of a battery of 11 subtests divided into a Verbal and Performance Scale. The Full Scale IQ score reflects overall, or global intelligence (Wechsler, 1967). Comprehensive administration instructions are included in the test manual. The test is appropriate for children aged 4 to 6 1/2 years. Appendix B includes a description of each of the WPPSI and the WPPSI-R subtests.

WPPSI-R

The WPPSI-R is the revised edition of the WPPSI, and is similar to its predecessor in content and format. The WPPSI-R consists of 12 subtests which are grouped into Verbal and Performance Scales. One subtest, Object Assembly, is new to the revised test. The age range of the WPPSI-R has been expanded to include children aged 3 to 7 1/2 years. Directions for standard administration of the instrument are included in the manual (Wechsler, 1989).

Procedure

Subjects were individually administered all subtests of the WPPSI and the WPPSI-R in counterbalanced order using standard administration procedures. Twelve subjects (6 male and 6 female) received the WPPSI first; the remaining 12 received the WPPSI-R first. The initial order in which the children were tested was determined randomly. The mean interval between the first and second testings was 30 days, with a range of 10 to 38 days. All tests were administered in the respective preschools during the normal hours of attendance in rooms set aside for individual testing. All testing was done by a female, Caucasian graduate student proficient in the administration of the tests. See Appendix D for the subject's WPPSI and WPPSI-R IQ scores.

Treatment of the Data

Group mean IQ scores were computed for the Verbal, Performance, and Full Scales of the WPPSI and the WPPSI-R, as well as for each of the subtests. Correlation coefficients were obtained for the WPPSI and WPPSI-R Verbal, Performance, Full Scales, and for the subtests common to the two instruments. Two tailed dependent *t* tests were computed to determine the comparability of the Verbal, Performance, and Full Scale IQs. The comparability of the individual subtests was determined

using the paired dependent *t* test. The required significance level was predetermined to be the .05 level.

CHAPTER IV

RESULTS AND DISCUSSION

The scores on the 11 parallel subtests and the 3 IQs derived from the WPPSI and the WPPSI-R were compared through Pearson *r* correlations and *t* tests of the paired mean differences for the total sample (*N* = 24). Descriptive statistics, *t* ratios, and correlations for the 14 comparable scores appear in Table 1.

Overall, results revealed consistently higher WPPSI IQs when compared to their WPPSI-R counterparts. The group mean WPPSI Verbal IQ score was 102.0 compared to a WPPSI-R Verbal IQ of 93.7 (*t* = 5.79). The WPPSI mean Performance IQ score was 108.6 compared to a WPPSI-R Performance IQ of 95.0 (*t* = 7.44). The mean WPPSI Full Scale IQ score was 105.83 compared to a WPPSI-R Full Scale IQ of 94.12 (*t* = 7.96).

Mean differences between the Verbal, Performance, and Full Scale IQs on the two tests were 8.3, 13.6, and 11.7, respectively, all in favor of the WPPSI. These differences are similar, although slightly larger in magnitude, to those found in the comparative study reported in the WPPSI-R manual (Wechsler, 1989), and in the Griggs (1991) and Milrod and Rescorla (1991) studies (see Appendix C). It is interesting to note that the Verbal IQ scores showed the least relative difference

between the two tests, thus supporting observations made by Flynn (1987) that scores on tests of verbal-educational abilities seem to be showing less increase in the population than do those of visuo-spatial abilities.

Comparison of the mean subtest scaled scores on the WPPSI with those on the WPPSI-R (see Table 1) indicates that, with the exception of two subtests, Sentences and Vocabulary, the WPPSI-R scaled scores are significantly lower than the scaled scores on the WPPSI. While this is certainly as would be predicted, the exceptions raise some interesting questions. In particular, the Vocabulary subtest on the WPPSI and the WPPSI-R yielded highly similar mean scaled scores (10.0 and 9.6) in the current study. A similar result was reported by Kaplan et al. (1991) when comparing WPPSI and WPPSI-R scores with a sample of "bright" children. These authors hypothesized that expressive language development, which the Vocabulary subtest measures, has not substantially increased in the American culture since the norming of the WPPSI 22 years ago. The results of the current study would seem to support this hypothesis with an African American population.

Table 1

WPPSI/WPPSI-R Comparisons of Scaled Scores for Total Sample (N=24)

	WPPSI		WPPSI-R		WPPSI/WPPSI-R	
Subtests	M	SD	M	SD	t	Pearson r
Information	9.54	2.55	8.50	2.67	2.87*	.77**
Vocabulary	10.04	3.08	9.66	2.79	.96	.79**
Arithmetic	10.28	2.37	8.70	1.96	4.04**	.66**
Similarities	11.20	3.77	9.54	3.06	3.58**	.79**
Comprehension	10.62	2.99	9.04	2.56	4.27**	.79**
Sentences	10.87	2.89	11.37	2.49	-.99	.59*
Animal House	12.54	3.46	11.58	3.29	2.79*	.87**
Pic. Completion	10.75	3.22	9.25	3.16	4.80**	.88**
Mazes	10.87	3.43	8.25	3.67	6.24**	.83**
Geo. Design	10.50	3.63	8.79	3.43	3.25*	.73**
Block Design	11.75	2.60	10.12	2.30	3.41*	.55*
Verbal IQ	102.00	15.16	93.70	12.08	5.79**	.90**
Performance IQ	108.62	18.54	95.00	16.62	7.44**	.87**
Full Scale IQ	105.83	18.09	94.12	16.69	7.96**	.91**

Note. *p< .01. **p< .001.

Correlation coefficients between the 3 IQs and the 11 comparable subtests were uniformly positive and significant (see Table 1). The Verbal IQs on the two tests correlated at the .90 level, the Performance IQs at the .87 level, and the Full Scale IQs at the .91 level. Individual subtests yielded correlation coefficients ranging from .55 (Block Design) to .88 (Picture Completion). These correlations are comparable to those reported in previous studies comparing the two instruments (Griggs, 1990; Wechsler, 1989) and support the view that the two tests are measuring the same construct.

Table 2 shows the range of absolute differences between WPPSI and WPPSI-R Verbal, Performance, and Full Scale IQs by varying magnitudes. Examination of the Table reveals that 88% of the sample attained Full Scale IQ scores 0 to 20 points higher on the WPPSI than on the WPPSI-R. None of the subjects had Verbal IQs on the WPPSI that exceeded their WPPSI-R scores by more than 20 points. However, Performance IQs were more variable; nearly one-fourth of the sample earned WPPSI Performance IQs that were 21 to 40 points higher than the comparable WPPSI-R Performance IQ. The greater variability in the observed differences between the WPPSI and the WPPSI-R Performance IQs may be due to the inclusion of a new subtest (Object

Assembly), to significant changes in the scoring of Geometric Design, and/or to other factors.

Table 2

Number of Subjects who Attained Varying Degrees of Difference between WPPSI and WPPSI-R Verbal, Performance, and Full Scale IQ Scores (N = 24)

	Number of IQ points higher on WPPSI than on WPPSI-R			
Scale	0-10	11-20	21-31	31-40
Verbal IQ	15 (62.5%)	9 (37.5%)		
Performance IQ	10 (42%)	9 (38%)	3 (13%)	2 (8%)
Full Scale IQ	11 (46%)	10 (42%)	3 (13%)	

CHAPTER V

SUMMARY

The publication of the WPPSI-R (Wechsler, 1989) has provided psychologists with a re-normed and updated version of the 1967 WPPSI. The revision of the WPPSI is timely. The last two decades have seen many cultural changes that specifically impact preschool children. For example, changes in family structure have led to increases in the number of children from all socioeconomic strata attending preschool. In addition, early detection and intervention programs such as Head Start have made educational services available to many young children in the lower socioeconomic status groups. Because these and other cultural changes are reflected in the performance of the WPPSI-R normative sample, clinicians assessing preschool children need data on the comparability of the original and revised Wechsler Preschool Scales.

Initial studies have suggested that scores on the WPPSI-R are consistently lower than those on the WPPSI (Griggs, 1991; Kaplan et al., 1991; Milrod & Rescorla, 1991; Wechsler, 1989). The WPPSI-R manual reported a 5 to 9 point difference in WPPSI/WPPSI-R mean IQ scores, however, the study did not stratify subjects by socioeconomic indices such as parental occupation or education level. Therefore, it was unclear whether

cultural changes had influenced intellectual performance on the WPPSI-R equally across the socioeconomic spectrum.

The present study compared the WPPSI and the WPPSI-R with a focus on African American children in the lower socioeconomic strata. WPPSI scores obtained were consistently higher than WPPSI-R scores (see Table 1). These findings, in combination with similar studies investigating between-test differences with high socioeconomic status youth (Milrod & Rescorla, 1991), suggest that indeed intellectual performance has increased across the entire socioeconomic spectrum. The mean WPPSI/WPPSI-R differences reflected in this study of 8, 13, and 11 points on the Verbal, Performance, and Full Scale IQs, respectively, add support to the hypothesis that changes in our society such as improved nutrition, and increased preschool attendance have contributed to building a generation of preschoolers who, as a group, perform better than did children of an earlier generation.

Overall, correlations between similar WPPSI and WPPSI-R subtests and IQs were high (see Table 1) and consistent with correlations reported for the standardization sample (Wechsler, 1989). These correlations are within the range that would be expected for two scales that assess the same constructs and similar abilities in young children.

The results of this study have a great deal of practical significance for educators and psychologists. The mean WPPSI/WPPSI-R IQ differences of 8 to 13 points found here indicate that use of the newer WPPSI-R could lead to an increase in the number of children classified as mentally deficient and placed in special education programming. For example, in the present study, four children who scored within the average range of intellectual ability on the WPPSI would be within the mentally deficient range based on their WPPSI-R scores (see Appendix D, numbers 1, 7, 20, and 24). This occurrence points to the necessity of using multiple sources of data and assessment techniques when making diagnostic decisions concerning a child's educational placement and programming.

A more complete analysis of the WPPSI/WPPSI-R IQ differences shown in Table 2 suggests that, for children similar to those studied here, Performance IQs on the WPPSI-R frequently may be more than 20 points lower than would be those attained on the WPPSI. Full Scale and Verbal IQs will likely also be lower on the revised scale, although the magnitude of difference may not be so extreme. Parents and educators need to be made aware of this eventuality, especially in cases where a child originally tested on the WPPSI is re-tested on the

WPPSI-R. What may appear to be a dramatic drop in IQ can be attributed, at least in part, to changes in the IQ test and norms being utilized for assessment.

It should be stressed that further research needs to e conducted with similar populations of children in order to determine the extent to which these data can be generalized.

REFERENCES

Anastasi, A. (1968). Psychological testing. New York: Macmillan.

Angoff, W. H. (1988). The nature-nurture debate, aptitudes, and group differences. American Psychologist, 43, 713-720.

Austin, J. J. (1970). The use of the WPPSI in early identification of mental retardation and preschool education. Proceedings of the Annual Convention of the American Psychological Association, 5, 913.

Berk, R. A. (Ed.). (1982). Handbook of methods for detecting test bias. Baltimore, MD: Johns Hopkins University Press.

Bracken, B. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. Journal of Psychoeducational Assessment, 4, 313-326.

Brooks, C. (1977). WISC, WISC-R, Stanford-Binet L.M., and WRAT: Relationships among children age 6 to 10 referred for psychological evaluation. Psychology in the Schools, 14, 30-33.

Carlson, L., & Reynolds, C. (1981). Factor structure and specific variance of the WPPSI subtests at six age levels. Psychology in the Schools, 18, 48-54.

Catron, D. W., & Catron, S. S. (1977). WISC-R vs. WISC: A comparison with educable mentally retarded children. Journal of School Psychology, 15, 264-266.

Clarizio, H. F. (1978). Nonbiased assessment of minority group children. Measurement and Evaluation in Guidance, 11, 106-113.

Cleary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. American Psychologist, 30, 15-41.

Crockett, B. K., Rardin, M. W., & Pasewark, R. A. (1976). Relationship of the WPPSI and subsequent Metropolitan Achievement Test scores in Head Start children. Psychology in the Schools, 13, 19-24.

Flynn, J. (1984). The mean IQ of Americans: Massive gains 1932 to 1897. Psychological Bulletin, 95, 29-51.

Flynn, J. (1987). Massive IQ gains in 14 Nations: What do IQ tests really measure? Psychological Bulletin, 101, 171-191.

Goldman, R., & Hartig, L. (1976). The WISC may not be a valid predictor of school performance for primary grade minority children. American Journal of Mental Deficiency, 80, 583-587.

Griggs, L. (1991). The Wechsler Preschool and Primary Scale of Intelligence and the Wechsler Preschool and Primary Scale of Intelligence-Revised: How Comparable Are the IQ Scores? Unpublished Masters Thesis, University of Northern Iowa.

Gutkin, T., & Reynolds, C. (1981). Factorial similarity for black and white children from the WISC standardization study. Journal of Educational Psychology, 73, 227-231.

Gutkin, T., & Reynolds, R. (1990). The handbook of school psychology, Second Edition. New York: John Wiley and Sons.

Gyurke, J., Stone, B., & Beyer, M. (1990). A confirmatory factor analysis of the WPPSI-R. Journal of Psychoeducational Assessment, 8, 15-21.

Hagin, R., Silver, A., & Corwin, C. (1971). Clinical-diagnostic use of the WPPSI in predicting learning disabilities in grade 1. Journal of Special Education, 5, 221-232.

Hartlage, L., & Steele, C. (1977). WISC and WISC-R correlates of academic achievement. Psychology in the Schools, 14, 15-18.

Hilliard, A. G. (1979). Standardization and cultural bias as impediments to the scientific study and validation of "intelligence". Journal of Research and Development in Education, 12, 47-58.

Hilliard, A. G. (1980). Cultural diversity and special education. Exceptional Children, 46, 584-588.

Kaplan, C., Fox, L., & Paxton, L. (1991). Bright children and the revised WPPSI: Concurrent validity. Journal of Psychoeducational Assessment, 9, 240-246.

Kaufman, A. (1973). Comparison of the performance of matched groups of black children and white children on the WPPSI. Journal of Consulting and Clinical Psychology, 41, 186-191.

Kaufman, A. (1990a). The WPPSI-R: You can't judge a test by its colors. Journal of School Psychology, 28, 387-394.

Kaufman, A. (1990b). Assessing adolescent and adult intelligence. Needham, MA: Allyn and Bacon.

Kaufman, A., & Hollenbeck, G. (1977). Comparative structure of the WPPSI for blacks and whites. Journal of Clinical Psychology, 30, 316-319.

Kaufman, A., & Kaufman, N. (1983). Manual for the Kaufman assessment battery for children. Circle Pines, MN: American Guidance Service.

Kutsick, K., Vance, B., Schwarting, F., & West, R. (1988). A comparison of three different measures of intelligence with preschool children identified at-risk. Psychology in the Schools, 25, 270-275.

Kutsick, K., & Wynn, E. (1988). Comparison of the K-ABC Achievement Scale and WPPSI IQs of preschool children. Psychological Reports, 63, 143-146.

Larry P. v. Riles, 495 F. Supp. 96 (N.D. Cal. 1979).

Lieblich, A., & Shinar, M. (1975). The predictive validity of WPPSI with Israeli children. Educational and Psychological Measurement, 35, 473-475.

Lippold, S., & Claiborn, J. (1983). Comparison of the WAIS and the WAIS-R. Journal of Consulting and Clinical Psychology, 51, 315.

Lowe, J., Anderson, H., Williams, A., & Currie, B. (1987). Long term predictive validity of the WPPSI and the WISC-R with Black preschool children. Personality and Individual Differences, 8, 551-559.

Lynn, R. (1990). The role of nutrition in secular increases in intelligence. Personality and Individual Differences, 11, 273-285.

Lynn, R., & Hampson, S. (1986). The rise of national intelligence: Evidence from Britain, Japan, and the U.S.A. Personality and Individual Differences, 7, 23-32.

Meile, F. (1979). Cultural bias in the WISC. Intelligence, 3, 149-164.

Milrod, R. J., & Rescorla, L. (1991). A comparison of the WPPSI and the WPPSI-R with high-IQ children. Journal of Psychoeducational Assessment, 9, 255-262.

Oakland, T., King, J., White, L., & Eckman, L. (1971). A comparison of performance on the WPPSI, WISC, and Stanford-Binet with preschool children. Journal of School Psychology, 9, 144-149.

Oldridge, O., & Allison, E. (1968). A review of the WPPSI. Journal of Educational Measurement, 5, 347-348.

Parents in Action on Special Education (PASE) v. Hannon, No. 74-C-3586 (N.D. Ill., July 16, 1980).

Partenio, I., & Taylor, R. (1985). The relationship of teacher ratings and IQ: A question of bias? School Psychology Review, 14, 79-83.

Quereshi, M., & McIntire, D. (1984). The comparability of the WISC, WISC-R, and WPPSI. Journal of Clinical Psychology, 40, 1036-1043.

Reschly, D. (1978). WISC-R factor structures among Anglos, Blacks, Chicanos, and Native American Papagos. Journal of Consulting and Clinical Psychology, 46, 417-422.

Reschly, D. (1981). Psychological testing in educational classification and testing. American Psychologist, 36, 1095-1111.

Reschly, D., & Reschly, J. (1979). Validity of the WISC-R factor scores in predicting achievement and attention for four sociocultural groups. Journal of School Psychology, 17, 355-361.

Reynolds, C. (1980). Differential construct validity of a preschool battery for Blacks, Whites, Males, and Females. Journal of School Psychology, 18, 112-125.

Reynolds, C., & Hartlage, L. (1979). Comparison of the WISC and the WISC-R regression lines for academic prediction with black and with white referred children. Journal of Consulting and Clinical Psychology, 47, 589-591.

Reynolds, C., Wright, D., & Dappen, L. (1981). A comparison of the criterion-related validity (academic achievement) of the WPPSI and the WISC-R. Psychology in the Schools, 18, 20-23.

Rosenbach, J., & Rusch, R. (1991). IQ and achievement: 1930s to 1980s. Psychology in the Schools, 28, 304-309.

Salvia, J., & Ysseldyke, J. (1988). Assessment. Boston, MA: Houghton-Mifflin.

Sandoval, J. (1979). The WISC-R and internal evidence of test bias with minority groups. Journal of Consulting and Clinical Psychology, 47, 919-927.

Sattler, J. (1988). Assessment of children. San Diego, CA: Sattler.

Teasdale, T., & Owen, W. (1989). Continuing secular increases in intelligence and a stable prevalence of high intelligence levels. Intelligence, 13, 255-262.

United States. (1991). An act to ammend the Individuals with Disabilities Education Act to strengthen such act, and for other purposes (Public Law 102-119). Washington, DC: U.S. G.P.O.

Walker, A. (Ed.). (1991). Thesaurus of psychological index terms. Arlington, VA: American Psychological Association.

Wasik, J., & Wasik, B. (1972). Use of the WPPSI and the WISC with culturally deprived children. Measurement and Evaluation in Guidance, 5, 280-285.

Wechsler, D. (1949). Wechsler intelligence scale for children. New York: The Psychological Corporation.

Wechsler, D. (1950). Cognitive, conative, and non-intellective intelligence. American Psychologist, 5, 78-83.

Wechsler, D. (1955). Wechsler adult intelligence scale. New York: Psychological Corporation.

Wechsler, D. (1967). Wechsler preschool and primary scale of intelligence. New York: The Psychological Corporation.

Wechsler, D. (1974). Wechsler intelligence scale for children-revised. New York: Psychological Corporation.

Wechsler, D. (1989). Wechsler preschool and primary scale of intelligence-revised. New York: Psychological Corporation.

Wechsler, D. (1991). Wechsler intelligence scale for children III. New York: Psychological Corporation.

White, D., & Jacobs, E. (1979). The prediction of first grade reading achievement from the WPPSI scores of preschool children. Psychology in the Schools, 16, 189-192.

APPENDIX A

INFORMED CONSENT

Dear Parents,

I am a UNI student doing a study comparing two IQ tests. I would like your permission to give your child each of these two tests while they are at preschool during the next few weeks. The IQ tests I am studying were made for preschool children and involve puzzles, drawings, and similar activities. Most children find the activities fun, but if your child doesn't want to do them then he/she doesn't have to. There is absolutely no penalty for not participating. I would be happy to show you the tests I will be using, talk to you more about my study, or discuss the results of the study with you. My phone number is 273-2321 (Cynthia, Department of Educational Psychology and Foundations). If you have any questions about the rights of research subjects, you may call the Graduate College, UNI, at 273-2748. Here are some things you should know about my study:

1. I am comparing two IQ tests, called the "WPPSI" and the "WPPSI-R", to see if Black children score higher on one than on the other.
2. My study will also look at the question of rising IQ scores in the United States.
3. There is no risk involved in your child taking these tests.
4. The results of your child's test will only be seen by me. All names will be replaced with numbers after testing is complete to ensure confidentiality.

Please sign below if you would like your child to participate, and thank you for helping me with my study.

Sincerely,

Cynthia Vandervelde

I am fully aware of the nature and extent of my participation in this project as stated above and any possible risks arising from it. I acknowledge that I have received a copy of this consent statement.

_____ _____

SIGN HERE Parent or Guardian date

_____

Child's name

_____

Researcher

APPENDIX B

BRIEF DESCRIPTIONS OF THE VERBAL AND PERFORMANCE SUBTESTS OF THE WPPSI AND THE WPPSI-R

Verbal Scale

Information

Consists primarily of questions about events or objects in the enviornment. For most questions, a brief verbal response is required. New to the WPPSI-R are several questions requiring the child to identify information by pointing to a picture. The subtest assesses the child's knowledge about the enviornment gained from experience rather than formal education.

Vocabulary

Requires the child to give a verbal definition for words presented orally. For the WPPSI-R, new items have been added that require the child to name a pictured object. The subtest assesses the child's knowledge gained from formal education and exposure to environmental variables.

Arithmetic

Begins with pictorial items, proceeds with simple counting items, and ends with word problems that measure arithmetic concepts. Full-color art work is new to the WPPSI-R. The subtest requires concentration and arithmetic reasoning skills.

Similarities

Includes both sentence completion items and items that require longer verbal descriptions of how two verbally presented objects or events are alike. New to the WPPSI-R are initial items requiring the child to point to a pictured item that shares a common feature with another set of pictured items. The subtest measures abstract reasoning skills.

Comprehension

Requires the child to give a verbal answer to questions that concern reasons or consequences for actions and events. The format is the same for the WPPSI and the WPPSI-R, however two-thirds of the items are new. The subtest requires logical reasoning skills as well as practical knowledge and social judgement.

Performance Scale

Sentences

Requires verbatim repetition of a sentence that the examiner reads aloud. New sentences were added to the WPPSI-R to extend the range upward and downward. The subtest measures immediate recall, and requires attention.

Animal House/Pegs

Placement of correct colored pegs in holes below a series of pictured animals. The subtest is changed only in name on the WPPSI-R. It is a measure of learning ability, and also requires manual dexterity, attention, and memory.

Picture Completion

Requires pointing to or naming what is missing in a picture of a common object or event. The WPPSI-R includes new art work and several new items. The subtest assesses visual discrimination skills, and requires reasoning skills and long term visual memory.

Mazes

Includes completion of paper and pencil mazes under time constraints. The WPPSI-R includes several new simple items to increase the downward range. The subtest assesses planning ability, perceptual organization skills, and memory.

Geometric Design

Requires the child to draw a copy of a design from a printed model. New to the WPPSI-R are several introductory items where the child points to one of four designs that matches the stimulus design. The subtest measures perceptual and visual-motor organization abilities.

Block Design

Includes reproduction of a pattern using flat red and white blocks. The range has been extended on the WPPSI-R to include some easier and some more difficult items. The subtest requires non-verbal problem solving abilities.

Object Assembly

This subtest is new to the WPPSI-R, and requires the child to assemble puzzle pieces within time constraints. The subtest measures non-verbal reasoning and visual-motor coordination skills.

Note. Adapted from Buckhalt, J. (1991). Test reviews: The Wechsler Preschool and Primary Scale of Intelligence-Revised. Journal of Psychoeducational Assessment, 9, 271-279.

APPENDIX C

IQ DIFFERENCES BETWEEN THE WPPSI AND WPPSI-R REPORTED IN COMPARATIVE STUDIES

	IQ		
SCALE	V	P	FS
STUDY			
Wechsler (1989)	5.0	9.0	8.0
Griggs (1991)	3.1	11.7	7.8
Milrod & Rescorla (1991)	3.5	9.2	6.0
Kaplan, Fox, & Paxton (1991)	6.7	9.7	8.4
Vandervelde (1992)	8.3	13.6	11.7

Note. WPPSI mean scores higher than WPPSI-R in all studies.

APPENDIX D

SUBJECTS' VERBAL, PERFORMANCE, AND FULL SCALE IQ SCORES ON THE WPPSI AND THE WPPSI-R

	TEST					
	WPPSI			WPPSI-R		
SCALE	V	P	FS	V	P	FS
SUBJECT						
1	105	105	106	86	81	81
2	110	122	117	97	91	94
3*	84	91	86	84	81	81
4*	77	78	76	77	70	70
5	105	93	99	92	86	88
6*	105	123	115	109	104	107
7	94	89	91	85	73	77
8*	105	115	111	98	116	107
9*	84	99	90	88	95	90
10*	112	115	115	105	102	103
11	105	100	103	94	93	93
12	94	104	99	85	93	87
13	75	82	76	73	77	72
14*	104	120	113	99	110	105
15	117	118	119	100	98	99
16*	137	155	152	129	134	139

	TEST					
	WPPSI			WPPSI-R		
SCALE	V	P	FS	V	P	FS
SUBJECT						
17*	104	115	110	87	104	94
18	120	127	126	111	116	115
19	124	129	129	121	122	125
20	85	101	92	81	79	78
21*	97	108	103	90	100	94
22*	116	134	127	99	97	98
23	89	81	84	86	74	78
24*	100	103	101	87	84	84

Note. * indicates WPPSI administered first.