THE EFFECTS OF CULTURAL BIAS: A COMPARISON OF
THE WISC-R AND THE WISC-III

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

Melissa Cox Ewing, B.S.

Denton, Texas

December, 1994
THE EFFECTS OF CULTURAL BIAS: A COMPARISON OF
THE WISC-R AND THE WISC-III

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

Melissa Cox Ewing, B.S.

Denton, Texas

December, 1994

It has been suggested that the use of standardized intelligence tests is biased against minorities. This study investigates the newly revised Wechsler Intelligence Scale for Children-III in which Wechsler states that the new scale has eliminated biased items. Comparisons of the scores on the WISC-R and the WISC-III of a clinical population of sixteen African American and eighteen Caucasian males, ages ten to sixteen, revealed significant differences between the two groups on the WISC-III. The minority scores decreased predictably from the WISC-R to the WISC-III, but the Caucasian scores increased rather than decreasing. The findings of this study do not support the predictions and goals of revision as stated in the manual of the WISC-III.
ACKNOWLEDGEMENT

All data for this study were obtained through the help and cooperation of the Ft. Worth Child Study Center and Dr. Harry Brick. Special thanks is extended to this Clinic and Dr. Brick for their help in the successful completion of this project.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>THESIS</td>
<td>1</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>29</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE 1</td>
<td>Demographic Variables for Racial Groups.</td>
<td>21</td>
</tr>
<tr>
<td>TABLE 2</td>
<td>Primary Diagnoses of Subjects.</td>
<td>22</td>
</tr>
<tr>
<td>TABLE 3</td>
<td>Ranges of IQ Scores for the Racial Groups.</td>
<td>22</td>
</tr>
<tr>
<td>TABLE 4</td>
<td>Means of Significant Differences of IQs Across Racial Groups.</td>
<td>23</td>
</tr>
<tr>
<td>TABLE 5</td>
<td>Means of Significant Differences of IQs Within Racial Groups.</td>
<td>23</td>
</tr>
<tr>
<td>TABLE 6</td>
<td>Analysis of Variance Table for Repeated Measures - Race by Test.</td>
<td>24</td>
</tr>
<tr>
<td>TABLE 7</td>
<td>Analysis of Covariance.</td>
<td>24</td>
</tr>
<tr>
<td>TABLE 8</td>
<td>Adjusted and Estimated Means.</td>
<td>24</td>
</tr>
<tr>
<td>TABLE 9</td>
<td>Analysis of Factor by Covariate Interaction.</td>
<td>24</td>
</tr>
<tr>
<td>TABLE 10</td>
<td>Scheffe' Post Hoc Analysis.</td>
<td>25</td>
</tr>
</tbody>
</table>
THE EFFECTS OF CULTURAL BIAS: A COMPARISON OF
THE WISC-R AND THE WISC-III

Early in the twentieth century, the United States experienced a surge in the
development of intelligence tests. The intelligence test as we know it was developed in
France by Binet and Simon for the purpose of assessing the educability of school children
(Edwards, 1971). Henry H. Goddard (1908) developed and revised this test in America,
using the intelligence scale to evaluate the abilities of the mentally retarded. The
development of intelligence tests introduced many new problems to the field of
psychology. One such problem was the debate over the exact definition of intelligence. It
seemed logical that the best approach to evaluating intelligence would be to establish a
precise definition of the term intelligence, and then create a test that would evaluate this
concept (Robb, Bernardoni, and Johnson, 1972).

However, throughout the history of intelligence testing there has been no
consensual definition for "intelligence" established (Samuda, 1975). This lack of consensus
has impacted the interpretation of intelligence test results and all issues associated with
intelligence testing. In particular, the ambiguous and varied definitions of "intelligence"
have influenced the issues surrounding the testing of American minorities, using
standardized intelligence tests (Samuda, 1975). It has been suggested that the use of
standardized intelligence tests is biased against minorities. Some researchers argue that
specific item analysis of standardized intelligence tests reveals little bias (Reynolds and
Rechsly, 1983), while others show evidence that the scores of minority children are significantly different from the scores of white children (Munford, 1978; Arinoldo, 1981; Munford and Munoz, 1980). It is the purpose of this study to determine if the new Wechsler Intelligence Scale for Children-III (WISC-III) has eliminated the cultural bias as claimed in the goals of its revision.

This issue of cultural bias, however, cannot be addressed until one first examines the many definitions of intelligence. Some general definitions of intelligence have been advocated which stress the ability of an individual to learn from the environment and from experience (Robb et al, 1972). For Binet, intelligence meant to comprehend, judge and reason well (Truch, 1989). He believed that the concept of intelligence was too complex and moved away from a broad definition of the term.

Other definitions include Terman's idea that intelligence is abstract thinking (Samuda, 1975). Burt suggests that intelligence is innate, general, cognitive ability (Samuda, 1975), while Wechsler believes that intelligence is the capacity to act purposefully, to think rationally and to deal effectively with the environment (Truch, 1989). He stressed the evaluation of global abilities. Vernon, on the other hand, emphasizes innate potentiality, good emotional adjustment, and appropriate educational stimulation (Robb et al, 1972).

All of these definitions encompass vague and undefinable terms, overemphasize the ability to reason abstractly, and tend to regard intelligence as an entity (Samuda, 1975). Wesman has advocated perhaps the least criticized definition which emphasizes intelligence as an attribute, not as an entity. He believes intelligence to be the summation
of learning experiences (Robb et al, 1972). His definition includes all items measured by intelligence, aptitude, and achievement tests and allows for the measurement of diverse abilities (Robb et al, 1972). Wesman suggests that it is important to consider not only innate abilities but all of the experiences during life that contribute to learning.

The definition by Wesman also incorporates the idea that "intelligent behavior is inescapably linked to and determined by the values and standards of society" (Samuda, 1975). This statement suggests that the discussion of intelligence as a sum of all learning experiences cannot be separated from the diverse cultural experiences of individuals in a society (Joseph, 1977).

Even before the development of tests of mental abilities, Galton had predicted the inferiority of some races to others (Joseph, 1977). Even Terman confirmed this presumed genetic inferiority of certain races with his revised scale in 1916. However, he seemed to ignore the results of various tests on his own scale which revealed that social class differences might account for the problem rather than genetic differences (Edwards, 1971). As the controversy widened, Arthur Jensen (1969) published an article advocating that individual differences on intelligence tests have a high degree of genetic influence (almost 80 percent). In addition, he suggests that any social-class differences in IQ were also linked to a genetic component (Loehlin, 1975). He believes that all of the evidence combined makes a reasonably sound hypothesis that genetic factors are strongly implicated in the intellectual differences between races (Jensen, 1969).

These debates were fueled by the age-old question of nature versus nurture. Both sides have collected vast amounts of evidence to support their beliefs about the nature of
intelligence. Some researchers have used twin and sibling studies, adoption studies, and foster care studies to investigate the relation of heredity and intelligence (Samuda, 1975). The advocates of the environmental side of the controversy believe that education and the social environment contribute to the development of intellect in an individual (Samuda, 1975).

Early investigators such as Galton and Goddard examined prominent families to arrive at the conclusion that heredity has the most influence on mental ability (Hothersall, 1990). Eysenck also believed that genetics accounts for the majority of difference in intelligence. He used many studies of twins and siblings to support his claims (Eysenck, 1971). Furthermore, Arthur Jensen advocated the idea of the heritability of intelligence (Jensen, 1969). He focused his attention on the differences between intelligence of different races, particularly blacks and whites (Jensen, 1980). His controversial article in 1969 clearly advocated that individual differences in IQ are due to genetic differences (Jensen, 1969).

The environmentalists, on the other hand, refute the arguments of hereditarians and suggest that intelligence is mainly influenced by environmental factors (Samuda, 1975). These individuals emphasize that unless children experience equal opportunities to learn what is required on intelligence tests, then differences in IQ scores will always exist (Good, 1954). In 1957, eighteen social scientists concluded in a special conference that there was no basis for innate racial difference in intelligence and that most differences of intelligence could be accounted for by environmental difference (Robb et al, 1972). David Wechsler wrote that he believed that intelligence tests were unfair to disadvantaged and
underprivileged minority groups (Samuda, 1975). However, he stated that it was not the IQ that made them unfair, but it was poor housing, poverty, and lack of opportunities that created inequality and bias within testing (Samuda, 1975). The largest criticism of testing of minority and deprived children has been that such standardized tests are unfair to children of different or deprived cultural backgrounds (Vernon, 1979). The original tests were standardized on the performance of white middle class children and conducted by white middle class psychologists. It was this fact that led to the controversy over the differences in IQ scores between minority and white children (Vernon, 1979). Within this controversy, the differences in IQ scores were said to be proof of the cultural bias that existed in standardized tests of intelligence. Some researchers argue that specific item analysis reveals little bias within standardized tests (Reynolds and Rechsly, 1983), while others show evidence that the scores of minority children are significantly different from the scores of white children (Munford, 1978; Arimoldo, 1981; Munford and Munoz, 1980). The term "bias" like the term "intelligence," however, is difficult to define. One researcher, Anne T. Cleary (1968) states that "a test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroup" being tested. In particular, these differences in scores are labeled "unfair" if they are consistently lower for one group when compared to another group (Cleary, 1968).

Many organizations such as the Association of Black Psychologists (ABP) have advocated and supported the belief that standardized intelligence tests are culturally biased against minority children (Samuda, 1975). They have sought and won several lawsuits
and court rulings on just this issue. For example, the ABP successfully prevented the use of such tests with black children in California through the court's ruling in the case of Larry P. et al. versus Wilson Riles et al. in 1972 (Samuda, 1975). This case held that the current use of standardized tests was "inappropriate and inadequate" because they were based on "white, middle-class norms, values and experiences" (Samuda, 1975). This case prevented the use of the current tests of intellectual ability for placement of minority children in special classes.

This case was particularly important in emphasizing the use of test data from standardized tests to place minority individuals in special classes. Many individuals felt that this test data was used to label and discriminate against minority students (Robb et al., 1972). These individuals believed that the differences of IQ scores between white children and minority children reinforced the attitudes of bias against those who belong to minority groups (Robb et al., 1972). Although these differences as discussed previously may be a result of environmental inequalities, many opponents of intelligence tests argue that teachers do not consider environmental differences but only consider test score differences (Robb et al., 1972).

Within the educational system, it is not uncommon for students to be stratified into ability groups with students of similar intellectual achievements (Samuda, 1975). Because standardized intelligence tests are typically used for grouping students, minority children will most often fall in lower groups as compared to white, upper-class students. This stratification system ranks children in terms of estimated potential (Samuda, 1975). This fact leads to lower self-concept and perceived lowered expectations of teachers for
children who are placed in the lower ability groups. Issues of equality and education have been emphasized since the Supreme Court's ruling in Plessy v. Ferguson (1896). This decision required the establishment of parallel systems of education and the separation of students in different schools (Plessy v. Ferguson, 1896). This principle of "separate but equal" was challenged in the case of Brown v. Board of Education of Topeka (1954). This case made it illegal to bar any student from attending a school on the basis of ethnicity only (Samuda, 1975).

This case attempted to eliminate racial segregation within the educational system. However, in the 1960's researchers suggested that the educational policy of ability grouping might be another form of racial segregation (Samuda, 1975). Heathers (1969) remarked that ability grouping may be an agency for maintaining and enhancing caste and class stratification in a society. The critical case of Hobson v. Hansen (1967) investigated the issues surrounding ability grouping. The court recognized that because tests are primarily standardized on white middle-class groups of students, they produce inaccurate and misleading test scores when given to lower class and minority students (Hobson v. Hansen, 1967). As a result, the students are classified according to their socio-economic or racial status rather than their ability to learn (Hobson v. Hansen, 1967).

More recently in 1975, Public Law 94-142 was passed which established that testing materials and procedures are to be selected and administered so as not to be racially or culturally discriminatory. PL 94-142 focuses on the education of handicapped children but is also intended to apply to all children including children with physical impairments and learning disabilities. This legislation also states that these children should
be educated with other non-handicapped children, and that special classes or separate schooling should occur only when the nature or severity of the handicap is such that education in regular classes with the use of supplementary aids and services cannot be achieved satisfactorily (PL 94-142, 1975). This legislation hopes to increase the equality of educational opportunities for all children and reduce the segregation of such children based on standardized test materials.

In addition to legislative actions, researchers have attempted to resolve the issues surrounding standardized tests of intelligence by developing culture-free or culture-fair tests. Cattell (1959) constructed perhaps the first culture-fair intelligence test. The purpose of this paper and pencil test was to provide a measure of ability which separates natural intelligence from that contaminated and obscured by education (Cattell and Cattell, 1959).

Another test was the Progressive Matrices Test which was developed in Great Britain by Raven (1965). This nonverbal test consists of 60 matrices or designs from each of which a part has been removed (Raven, 1965). The Leiter International Performance Scale is another nonverbal test in which both the examiner and examinee perform rather than verbalize questions and answers (Leiter, 1969).

These tests and others were attempts by researchers to overcome the controversy surrounding traditional intelligence tests. These culture-fair tests, however, have been criticized because they seem to measure only a very limited aspect of intelligence and are not "culturally fair" to more than a few isolated cultures (Robb et al, 1972). The tests
often fail to give much useful information about the individual being tested and often do not give adequate information on reliability and validity (Robb et al, 1972).

The culture-fair intelligence tests seemed to be a practical solution to the controversy surrounding traditional intelligence tests, but to date, they have not succeeded in replacing the traditional standardized intelligence tests. However, test developers continue to be concerned with issues of test bias and culturally fair intelligence tests. One such test developer is David Wechsler. In the 1930's, he combined intellectual tasks developed by others in the field to create his clinical test battery (Kaufman, 1990). His original Wechsler-Bellevue Form I was eventually extended downward to create the Wechsler-Bellevue Form II which covered ages five to fifteen instead of ten to fifty-nine (Kaufman, 1990). This second form eventually produced the successful Wechsler Intelligence Scale for Children (WISC) (Wechsler, 1949).

Twenty-five years later Wechsler (1974) published the Wechsler Intelligence Scale for Children-Revised (WISC-R). Since the WISC was a downward extension of an adult version, the WISC-R added simpler items which were more suitable for children (Sattler, 1988). It covered an age range from 6-0 to 16-11 years and contained twelve different subtests such as information, arithmetic, vocabulary, digit span, block design and mazes (Wechsler, 1974). A total of seventy-two percent of the WISC items were retained in the WISC-R with sixty-four percent intact and eight percent substantially changed (Sattler, 1988). In addition, the WISC-R was standardized on a stratified or representative sample of the population of the United States. Unlike the WISC which used only white children, the WISC-R included minority groups as well (Sattler, 1988).
Due to continuing controversies over test bias and the failure of culture-fair intelligence tests to replace traditional intelligence tests, Wechsler attempted to minimize the bias in his tests by continuing to revise them. He stated that some of the goals of his recently published WISC-III were to:

1. periodically update norms,
2. maintain the basic structure and content of the WISC-R,
3. enhance the factor structure underlying the WISC-R, and
4. improve various subtests.

"One of the focal points of subtest improvement was the minimization of bias in content" (Wechsler, 1991). "An extensive empirical analysis of the WISC-R normative data was conducted. Several methods (Angoff, 1982; Reynolds, Willson, & Jensen, 1984; Wright & Stone, 1979) were then used to examine item-bias statistics for gender, ethnic, and regional bias" (Wechsler, 1991). He claimed that the very few items that were found to be biased were replaced in the WISC-III (Wechsler, 1991). Although no data were available within the manual, Wechsler stated that item-bias analyses were later conducted with the WISC-III "tryout data" (Wechsler, 1991). Furthermore, he tested an oversample of 400 minority children in addition to the standardization sample and analyzed their scores for bias (Wechsler, 1991).

Finally, Wechsler claims that he used reviewers to examine items for bias. One panel was composed of psychologists "familiar with the ethnic-bias studies of the WISC-R" (Wechsler, 1991). "This group provided written evaluations of all WISC-R items and
the proposed WISC-III items" (Wechsler, 1991), but this information was not available within the manual.

Reynolds and Rechsly (1983) support the claims of Wechsler that his tests contain very little bias. These researchers looked at four sociocultural groups and found that no item bias existed for Anglos, Blacks, or Hispanics. Likewise, Sandoval (1979) examined the WISC-R for evidence of cultural bias. The performance of White, Black, and Hispanic children were assessed on the internal criteria of reliability and a number of order-of-item-difficulty measures. Sandoval (1979) found that the minority children did not differ from the Anglo-American children, and there was no clear pattern of items that were more difficult for the minority children. Therefore, he concluded that the WISC-R was not biased for minority children.

In addition, Vance, Hankins, and McGee (1979) found evidence that actually showed unexpected differences between minority children and white children. They found that the minority children performed better on verbal tasks than the white children. However, they suggested possible reasons for their unusual findings which included the high degree of homogeneity of their sample. Although they suggest that their findings be interpreted cautiously, the results do suggest that the WISC-R may be unbiased and that some minority children can actually perform better than white children.

On the other hand, Reynolds and Gutkin (1981) looked at the black and white intelligence difference in their study of 285 matched pairs of black and white children. Their analysis of the results on the WISC-R revealed that whites exceeded blacks on all tasks except digit span—a test of rote memory (Reynolds et. al, 1981).
Munford (1978) also reported that, despite earlier claims by Wechsler that the WISC-R was improved in regard to bias, cultural bias was still evident in the scores of black children on specific subtests and overall IQ. Munford and Munoz (1980) studied Hispanic children, finding the same differences in IQ and concluding that the WISC-R was as culturally biased as the WISC. Arinoldo (1981) also found that black and white children differed significantly in their Wechsler Full Scale IQs. He found these differences in preschool and school-age children.

In another study, Munford, Meyerowitz, and Munford (1980) compared the results of black and white children on the WISC and WISC-R. They found no differences for the whites between their scores on the WISC and WISC-R. However, they found that the blacks scored significantly lower on the WISC-R than on the WISC on all three IQ measures. They concluded that their results indicate increased cultural bias, despite efforts to produce less bias.

Presumably, an unbiased intelligence test should provide accurate data on the assessment of intelligence in children of all cultural backgrounds. However, researchers have still not reached a consensus on the controversy of test bias that surrounds traditional intelligence tests. The primary concern as Munford (1978) pointed out is that differences in intelligence scores may cause black or minority children to receive developmental disability labels. These labels coming from standardized test scores would lead to inappropriate assumptions of the abilities of minority children.

Likewise, tests that produce lower IQs for certain minority children may also support the hypothesis that these low IQ scores will predict poor school performance.
(Cleary, Humphreys, Kendrick and Wesman, 1975). Therefore, more black children than white children would presumably fail to achieve academically and most likely never attain the credentials required by a higher occupational status (Scarr, 1981). This in turn might lead these black children to miss the social prestige and economic security that a higher occupational status provides (Scarr, 1981).

However, attempts to develop a culturally fair test have not been successful (Sattler, 1988) and, in fact, may not be possible (Loehlin, 1975). Williams (1972), for example, developed a Black Intelligence test which included items taken from black culture. Research showed that this test had questionable validity and had low correlation with a currently accepted achievement test. This Black intelligence test was useful in building Black pride but was not a good predictor of cognitive abilities or achievement in school (Sattler, 1988). In fact, the WISC-R appeared to be a better predictor of achievement in black children than this newly developed test.

Myra Shimberg (1929) attempted to see if a test could be created to reverse test bias. She standardized an "Information test A" on urban children and an "Information test B" on rural children. The two tests were then given in both rural and urban schools. The results showed that the urban children did better on test A and the rural children did better on test B. Therefore, she concluded that the earlier tests that found cultural biases may have been finding problems with the tools of measurement and not actual intellectual differences (Shimberg, 1929).

Furthermore, even attempts to write an intelligence test in the natural dialect of the black children did not appear to make these test more culturally fair. Quay (1971)
translated the Stanford-Binet into Black English and found that black inner-city children performed no better on this test than one written in Standard English. Although Quay did not account for nonlinguistic factors that might have influenced the results, this study does suggest that some tests may possess a bias that cannot be eliminated.

Most likely no test will be developed that can completely eliminate the influence of cultural learning. It appears that all human experience is influenced by culture. Scarr (1981) observed: "Intelligence tests are not tests of intelligence in some abstract, culture-free way. They are measures of the ability to function intellectually by virtue of knowledge and skills in the culture of which they [are a] sample." In other words, it may be more appropriate to study the Wechsler Intelligence Scale for Children-III by looking at whether the improvements have reduced the culturally loaded materials within the test that create the differences between minority and white children. It may be that several factors such as cultural biases, environmental factors, and genetic differences interact to reflect the observed differences between various ethnic groups.

A few researchers have begun to investigate the WISC-III. Kamphaus and Platt (1992) computed the new subtest specificities for the WISC-III using a method previously described in 1979. Kramer (1993) provided information to guide in the interpretation of the individual subtests within the WISC-III. He included tables that can be used as the psychometric guidelines for interpretation. While these researchers investigated specific areas within the new Wechsler scale, Sattler and Atkinson (1993) compared the WISC-III with the WPPSI-R. These researchers wanted to assess the "robustness of items across the WPPSI-R and WISC-III in terms of test-item equivalence" (Sattler et al, 1993). They
found evidence to suggest that the validity of the WPPSI-R and WISC-III extends beyond IQ to the level of the item.

Smith, Stovall, and Geraghty (1994) compared the WISC-R and the WISC-III for use in re-evaluations of mentally retarded individuals. They found that in most cases IQ scores decreased from the WISC-R to the WISC-III. However, they found the two scales to be highly related and the decreases to be consistent with the average decrease in scores reported in the WISC-III manual. They conclude that the use of the WISC-III is strongly supported; however, "for states with rigid criteria for placement in learning disabilities programs that require an IQ score of 90 or above, this difference in scores could be problematic" (Smith et al, 1994). Nagle and Daley (1994) also compared the WISC-R and WISC-III in their study of the longitudinal comparability of the WISC-R and WISC-III in EMR students. They found five to eight point declines in IQ from the WISC-R to the WISC-III. Although they emphasize the need for further research, they suggest that these declines in IQ can be expected and may be attributed to "changing mental growth patterns as well as the general phenomena of lower scores obtained on newly revised and restandardized tests of intelligence" (Nagle & Daley, 1994).

The present study also compares the WISC-R to the WISC-III. This study attempts to determine if Wechsler has successfully reduced some of the culturally loaded items that he claimed to have found and eliminated in the WISC-III (Wechsler, 1991).

Hypothesis 1: Verbal, Performance and Full Scale IQ scores will be lower on the WISC-III than the WISC-R. Hypothesis 2: African American males will score significantly lower than Caucasian males on the WISC-R for Verbal, Performance, and Full Scale IQ
scores. Hypothesis 3: There will be no differences between African American males and Caucasian males on the WISC-III.

If Wechsler has indeed improved and reduced the culturally loaded items in the WISC-III, then the WISC-R and WISC-III scores of the white children should remain relatively the same with only an approximate five, two, and seven point decrease on the Full Scale IQ, Verbal IQ, and Performance IQ, respectively. These decreases are reported by Wechsler to be the average decrease in scores typically found when comparing the older intelligence test, the WISC-R, with the new updated intelligence test, the WISC-III. However, the scores among the minority children should reveal increases that rise above the typical decrease predictable in the difference between IQs on the WISC-R and the WISC-III.

Methods

Subjects:

The subjects include eighteen Caucasian males and sixteen African American males between the ages of ten and sixteen (Mean age 13.222 and 13.000 for Caucasian and African American males, respectively). Means of various demographic variables and the range of IQ scores for these groups are found in Tables 1 and 3, respectively. All of the subjects were clients at the Ft. Worth Child Study Center. These youth were receiving psychological services primarily for Attention-Deficit Hyperactivity Disorder and specific developmental disorders (see Table 2) such as Developmental Reading and Developmental Arithmetic Disorders. Children with Pervasive Developmental Disorders, Autism,
Psychotic Disorders and Major Clinical Depression were excluded from inclusion in this study.

**Instrumentation:**

All of the subjects had been administered the WISC-R prior to administration of the WISC-III. Due to the fact that the WISC-III includes one new subtest, this subtest score (Symbol Search) was dropped from the sum of scores. The overall IQ scores on the WISC-III were based on the same set of ten subtests as the WISC-R. These subtests include Picture Completion, Information, Coding, Similarities, Picture Arrangement, Arithmetic, Block Design, Vocabulary, Object Assembly, and Comprehension.

Wechsler (1991) reports the standard error of measurement (SEM) for the WISC-III subtests, IQ scales, and factor-based scales. The SEM provides an estimate of the amount of error in an individual's observed test score. The greater the reliability, the less the SEM (Wechsler, 1991). The stability of scores on the WISC-III was also studied by analyzing the scores of 353 children who were tested twice (Wechsler, 1991). The WISC-III possesses adequate stability across time and across age groups with an approximate increase of seven to eight points in the Full Scale IQ score over a short retest period (Wechsler, 1991).

Wechsler (1991) reports that the WISC-R and WISC-III are highly correlated with .90 and .89 for Verbal IQ and Full Scale IQ, respectively. The correlation between Performance IQ is slightly lower (r=.81).

A comparison of the mean Full Scale IQ score shows that the WISC-III FSIQ is approximately five points less than the WISC-R FSIQ. The WISC-III VIQ and PIQ
scores are two and seven points less than corresponding WISC-R scores. In addition, Wechsler (1991) found similar results with a clinical sample of children.

Analysis:

Means and standard deviations were computed separately for the two groups (Caucasian and African American) on the demographic variables (age, income, mother's highest educational level, and father's highest educational level), the Verbal, Performance, and Full Scale IQs. For the WISC-R and WISC-III, the three IQ scales have a mean of 100 and a standard deviation of 15.

T-tests using separate variance estimates were employed to determine any significant differences among the demographic variables, WISC-R IQ scores and WISC-III IQ scores across the two racial groups. Paired t-tests were also employed to compare the difference between the WISC-R and WISC-III IQ scores within each racial group. The data was also subjected to an analysis of variance with repeated measures. An analysis of covariance and post hoc analysis were also employed to investigate the significance of the differences.

Results

In comparison of demographic variables between the two racial groups, no significant differences were found for age of subjects, parent's income, mother's educational level, and father's educational level (see Table 1). Table 4 presents data from the independent t-tests for the IQs that significantly differed between groups on the WISC-R and the WISC-III. On the WISC-R, the Verbal IQ was lower for African
American males than for Caucasian males. On the WISC-III, the Verbal, Performance and Full Scale IQs were all lower for African American males than for Caucasian males.

Table 5 presents data from the paired t-tests for the IQs within the racial groups. Significant differences between the two tests, WISC-R and WISC-III, were only found within one racial group, African American males. Verbal, Performance, and Full Scale IQs all decreased from the WISC-R to the WISC-III.

Table 6 presents the significant F ratios found in analysis of variance with repeated measures. The data represents the separate main effects of race and test ($F_{1,32} = 7.93, p < .01; F_{1,32} = 5.72, p < .05$, respectively) as well as the interaction effects of race and test ($F_{1,32} = 6.73, p < .05$). This table reveals that racial differences interacted with the two tests, creating significant differences between the two groups on the tests.

Table 7, 8, and 9 presents the significant F ratios found in the analysis of covariance with WISC-III Full Scale IQs as the dependent variable, WISC-R Full Scale IQs as the covariate and race as the factor. The first portion of the data in Table 7 represents a one-way analysis of variance for the WISC-III ($F_{1,32} = 10.78, p < .01$). This data reveal that a significant difference exists between the two groups on the WISC-III. The second portion of the data in Table 7 represents the analysis of covariance including the analysis of variance and regression statistics associated with the covariate, WISC-R ($F_{1,31} = 6.99, p < .05$). A significant difference continues to exist between the two groups on the WISC-III, even after taking into account the original differences between the two groups on the WISC-R.
The data in Table 8 display the observed means, the adjusted means, which are adjusted for the covariate, and the estimated means, which are the cell means estimated with knowledge of the factor of race. Finally, the data in Table 9 represent the test of the factor by covariate interaction term. This tests whether the parallel slopes assumption is warranted. Since this interaction is not significant ($F = .04$, sig. of $F = 0.849$), the assumption of parallel slopes is not rejected. In other words, the effect of change in the covariate, WISC-R scores, on the dependent variable, WISC-III scores, is the same across both groups of the factor, race. In other words, the original differences between the two groups on the WISC-R does not impact the significant difference found between the two groups on the WISC-III. Any changes in the WISC-R scores similarly effects the WISC-III scores across both racial groups.

Table 10 displays the results of the Scheffe' post hoc analysis. A comparison of the means reveals a significant effect for group at the .01 level and a significant effect for test and group by test interaction at the .05 level.

Discussion

Continuing controversy over test bias and the failure of culturally-fair intelligence tests to replace traditional intelligence tests has led researchers to continue to investigate the differences between IQs for minority and Caucasian individuals. The present study analyzed results from the testing of sixteen African American males and eighteen caucasian males on both the WISC-R and WISC-III. It was hypothesized that IQ scores would decrease from the WISC-R to the WISC-III. Wechsler reported an average
decrease in scores of five, two, and seven points for the full scale, verbal, and performance IQ scores with WISC-III scores lower than WISC-R scores.

In this study, predicted decreases were found for the African American males between the WISC-R and the WISC-III. Although significant differences were found within the racial group of African American males between the two tests, these differences become nonsignificant upon allowing for the five, two, and seven point decrease in IQ scores (Full Scale, Verbal, and Performance, respectively) from the WISC-R to the WISC-III as reported by Wechsler. However, these same decreases should have been found on the Caucasian males as well. The present data reveal no five, two, or seven point decreases (respectively) within the IQ scores of the Caucasian males. In fact, Full Scale IQ and Performance IQ mean scores for Caucasian males increased from the WISC-R to the WISC-III. Verbal IQ mean scores decreased only slightly from the WISC-R to the WISC-III (Diff. = 0.61).

Furthermore, upon examination of past research on the WISC-R, it was hypothesized that African American males would score significantly lower on the WISC-R than Caucasian males. However, the scores of the African American males were only significantly lower than the Caucasian males on Verbal IQ. Performance and Full Scale IQ scores on the WISC-R were not significantly different for the two racial groups.

The final hypothesis stated that no differences would be found between the two racial groups on the Verbal, Performance, or Full Scale IQ scores of the WISC-III. This hypothesis was based on the research by Wechsler and the goals of revision presented in the WISC-III manual that suggested that the WISC-III had been improved and culturally-
Table 1
Demographic Variable for Racial Groups

<table>
<thead>
<tr>
<th>Race</th>
<th>Variable</th>
<th>Min/Max</th>
<th>Mean</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>Age (yrs)</td>
<td>10/16</td>
<td>13.222</td>
<td></td>
</tr>
<tr>
<td>African Am.</td>
<td>Age (yrs)</td>
<td>11/16</td>
<td>13</td>
<td>0.222</td>
</tr>
<tr>
<td>Caucasian</td>
<td>Income ($)</td>
<td>2500/76000</td>
<td>27686</td>
<td></td>
</tr>
<tr>
<td>African Am.</td>
<td>Income ($)</td>
<td>2500/60000</td>
<td>19146</td>
<td>8540.000</td>
</tr>
<tr>
<td>Caucasian</td>
<td>Mom's Educ.</td>
<td>9/16</td>
<td>13.111</td>
<td></td>
</tr>
<tr>
<td>African Am.</td>
<td>Mom's Educ.</td>
<td>6/20</td>
<td>11.813</td>
<td>1.298</td>
</tr>
<tr>
<td>Caucasian</td>
<td>Dad's Educ.</td>
<td>11/16</td>
<td>13.8</td>
<td></td>
</tr>
<tr>
<td>African Am.</td>
<td>Dad's Educ.</td>
<td>10/18</td>
<td>12.455</td>
<td>1.345</td>
</tr>
</tbody>
</table>

* p < .05
** p < .01

Table 2
Primary Diagnoses of Subjects

<table>
<thead>
<tr>
<th>Racial Group</th>
<th>Number in Group</th>
<th>Diagnosis</th>
<th>Percentage with Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>18</td>
<td>Attention Deficit-Hyperactivity Disorder ADHD</td>
<td>89.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parent-Child Problem</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Academic Problem</td>
<td>5.5</td>
</tr>
<tr>
<td>African Am.</td>
<td>16</td>
<td>ADHD</td>
<td>62.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Borderline Intellectual Functioning</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Academic Problem</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conduct Disorder</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Psychological factors Affecting</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Physical Conditions</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3
Ranges of IQ Scores for the Racial Groups

<table>
<thead>
<tr>
<th>Racial Group</th>
<th>Test/IQ</th>
<th>Min/Max</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>WISC-R/VIQ</td>
<td>57/128</td>
<td>97.72</td>
<td>98.5</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>WISC-R/PIQ</td>
<td>51/128</td>
<td>98.78</td>
<td>100</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>WISC-R/FSIQ</td>
<td>50/116</td>
<td>97.94</td>
<td>99.5</td>
<td>92</td>
</tr>
<tr>
<td>African Am.</td>
<td>WISC-R/VIQ</td>
<td>65/106</td>
<td>87.06</td>
<td>86</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>WISC-R/PIQ</td>
<td>54/118</td>
<td>91.69</td>
<td>92.5</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>WISC-R/FSIQ</td>
<td>59/110</td>
<td>88.06</td>
<td>88</td>
<td>59</td>
</tr>
<tr>
<td>Caucasian</td>
<td>WISC-III/VIQ</td>
<td>63/123</td>
<td>97.11</td>
<td>92.5</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>WISC-III/PIQ</td>
<td>60/125</td>
<td>110.2</td>
<td>104</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>WISC-III/FSIQ</td>
<td>58/123</td>
<td>98.28</td>
<td>100</td>
<td>119</td>
</tr>
<tr>
<td>African Am.</td>
<td>WISC-III/VIQ</td>
<td>62/108</td>
<td>80.56</td>
<td>80.5</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>WISC-III/PIQ</td>
<td>54/113</td>
<td>82.44</td>
<td>81.5</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>WISC-III/FSIQ</td>
<td>56/111</td>
<td>79.87</td>
<td>80.5</td>
<td>78</td>
</tr>
</tbody>
</table>

### Table 4
Means of Significant Differences of IQs Across Racial Groups

<table>
<thead>
<tr>
<th>Test/ IQ</th>
<th>African Am</th>
<th>Caucasian</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>WISC-R/VIQ</td>
<td>87.06</td>
<td>97.72</td>
<td>10.66*</td>
</tr>
<tr>
<td>WISC-III/VIQ</td>
<td>80.56</td>
<td>97.11</td>
<td>16.55**</td>
</tr>
<tr>
<td>WISC-III/PIQ</td>
<td>82.44</td>
<td>100.22</td>
<td>17.78**</td>
</tr>
<tr>
<td>WISC-III/FSIQ</td>
<td>79.87</td>
<td>98.28</td>
<td>18.41**</td>
</tr>
</tbody>
</table>

* $p < .05$
** $p < .01$
Table 5
Means of Significant Differences of IQs within Racial Groups

<table>
<thead>
<tr>
<th>Group/IQ</th>
<th>WISC-R</th>
<th>WISC-III</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>African Am./ VIQ</td>
<td>87.06</td>
<td>80.56</td>
<td>-6.50*</td>
</tr>
<tr>
<td>African Am./ PIQ</td>
<td>91.69</td>
<td>82.44</td>
<td>-9.25**</td>
</tr>
<tr>
<td>African Am./ FSIQ</td>
<td>88.06</td>
<td>79.87</td>
<td>-8.19**</td>
</tr>
<tr>
<td>Caucasian/ VIQ</td>
<td>97.72</td>
<td>97.11</td>
<td>-0.61</td>
</tr>
<tr>
<td>Caucasian/ PIQ</td>
<td>98.78</td>
<td>100.22</td>
<td>1.44</td>
</tr>
<tr>
<td>Caucasian/ FSIQ</td>
<td>97.94</td>
<td>98.28</td>
<td>0.33</td>
</tr>
</tbody>
</table>

* p < .05
** p < .01

Table 6
Analysis of Variance Table for Repeated Measures
Race by Test

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td>3388.34</td>
<td>1</td>
<td>7.93**</td>
</tr>
<tr>
<td>Test</td>
<td>261.27</td>
<td>1</td>
<td>5.72*</td>
</tr>
<tr>
<td>Race by Test</td>
<td>307.5</td>
<td>1</td>
<td>6.73*</td>
</tr>
</tbody>
</table>

* p < .05
** p < .01
Table 7
Analysis of Covariance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td>2868.67</td>
<td>1</td>
<td>10.78**</td>
</tr>
</tbody>
</table>

Analysis of Variance & Regression Associated with Covariate

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>5631.29</td>
<td>1</td>
<td>60.53**</td>
</tr>
<tr>
<td>Race</td>
<td>650.08</td>
<td>1</td>
<td>6.99*</td>
</tr>
</tbody>
</table>

* \( p < .05 \)
** \( p < .01 \)

Table 8
Adjusted and Estimated Means

<table>
<thead>
<tr>
<th>Group</th>
<th>Observed Means</th>
<th>Adjusted Means</th>
<th>Estimated Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>98.27778</td>
<td>93.72206</td>
<td>98.27778</td>
</tr>
<tr>
<td>African American</td>
<td>79.87500</td>
<td>84.43071</td>
<td>79.87500</td>
</tr>
</tbody>
</table>

Table 9
Analysis of Factor by Covariate Interaction

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariate: WISC-R</td>
<td>5447.68</td>
<td>1</td>
<td>56.73**</td>
</tr>
<tr>
<td>Factor: Race</td>
<td>34.03</td>
<td>1</td>
<td>0.35</td>
</tr>
<tr>
<td>Factor by Covariate</td>
<td>3.55</td>
<td>1</td>
<td>0.04</td>
</tr>
</tbody>
</table>

* \( p < .05 \)
** \( p < .01 \)
Table 10
Scheffé Post Hoc Analysis

<table>
<thead>
<tr>
<th>Source</th>
<th>t-Value</th>
<th>Sig. t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>2.81561</td>
<td>0.00827**</td>
</tr>
<tr>
<td>Test</td>
<td>2.39117</td>
<td>0.02285</td>
</tr>
<tr>
<td>Group by Test</td>
<td>-2.59414</td>
<td>0.01419</td>
</tr>
</tbody>
</table>

*p < .05
**p < .01

loaded items within the test removed to provide an intelligence test that more adequately evaluated children of any race or ethnicity. This study found that there was a significant difference between the Verbal, Performance, and Full Scale IQ scores of the two racial groups. African American males scored significantly lower than Caucasian males. Although the decreases from the WISC-R to the WISC-III followed the predicted trend for African American males, the scores of Caucasian males did not decrease as predicted. The difference in scores on the WISC-III between the two groups was accented by the unpredictable increase of Performance and Full Scale IQ scores for the Caucasian males. This difference suggests an ultimate difference in the two racial groups that was also supported by the results from the analysis of variance for repeated measures in Table 3. Controlling for initial differences between the two groups on the WISC-R using an analysis of covariance decreased the discrepancy between the groups on the WISC-III ($F = 10.78, p < .01$ to $F = 6.99, p < .05$). However, differences between the two racial groups on the WISC-III Full Scale IQ scores were still significant at the .05 level. Post
hoc comparisons of the means also found a significant difference beyond the .05 level for test, group, and group by test interaction.

These facts suggest that there was indeed a difference in performance of the two racial groups on the WISC-III. While African American scores decreased in the direction predicted by Wechsler, the Caucasian scores did not. Controlling for initial differences between the two groups did not eliminate the significant differences between the two groups on the WISC-III.

These findings do not support the predicted decreases reported by Wechsler between the WISC-R and the WISC-III. Significant differences on the performance of the two racial groups on the WISC-III do not support the claims of Wechsler and the goals of revision within the WISC-III manual. Differences between African American males and Caucasian males still exist on the three IQ scores for the WISC-III.

Within this study there are a few factors that should be considered when examining the differences found by this data. The sample size of both groups is small and further study might be suggested with a larger number of individuals. These results also only investigated the differences between males of the two racial groups because data for this gender was more readily available in this clinical population. Further research should compare the differences among female subjects and the differences between male and female subjects.

Furthermore, the information in this study was obtained through retrieval of archival data from the files of the subjects included in the study. Therefore, the subjects could not be administered the two tests in a counterbalanced design. All subjects were
administered the WISC-R prior to the WISC-III. The subjects were also all part of a clinical sample, and comparisons were not made on individuals outside of this clinical population.

Despite these specific factors, there appears to be a significant difference between the WISC-III Verbal, Performance and Full Scale IQ scores for African American and Caucasian males. African American males had significantly lower IQ scores on the WISC-III in comparison to the Caucasian males. The scores of Caucasian males revealed no differences between the WISC-R and the WISC-III, while African American males decreased the predicted amount from the WISC-R to the WISC-III. The findings of this study do not support the predictions and goals of revision as stated in the manual of the Wechsler Intelligence Scale for Children--Third Edition.
REFERENCES


**COURT CASES**


