AN ANALYSIS OF THE PERFORMANCE OF A CLINICAL SAMPLE OF AFRICAN AMERICAN, CAUCASIAN, AND HISPANIC CHILDREN ON THE WISC-III

DISSERTATION

Presented to the Graduate Council of the

University of North Texas in Partial

Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Melissa L. Ewing, B.S., M.S.

Denton, Texas

December, 1998

Ewing, B.S., M.S., Melissa L., An Analysis of the Performance of a Clinical Sample of African American, Caucasian, and Hispanic Children on the WISC-III. Doctor of Philosophy (Clinical Psychology), December, 1998, 90 pp., 15 tables, references, 99 titles.

The goals of revision for the Wechsler Intelligence Scale for Children-Third Edition included enhancement of the factor structure, improvement of subtests, and revision of norms. The researchers reported that the very few items that were found to be biased were replaced. The WISC-III performance of a clinical sample of African American, Caucasian, and Hispanic children was analyzed to determine if the test bias was eliminated as claimed in the goals of the revision. Demographic information and test scores were obtained for 241 children, males and females, between the ages of seven and sixteen. All of the subjects were clients at a metropolitan children's clinic, receiving psychological services primarily for Attention-Deficit Hyperactivity Disorder.

A multivariate analysis of variance indicated statistically significant differences among the groups on the three IQ scores and all subtests excluding digit span and coding. A comparison of the demographic variables among the three racial groups produced a mixture of results. Significant differences were found among the groups on the variables of income, father's education, and mother's education. When these variables were individually factored out using a multivariate analysis of covariance, significant differences were still present among the three ethnic groups on the subtests and IQ scores. However,

when the Full Scale IQ was held constant, no significant differences were found among the three ethnic groups on Verbal IQ, Performance IQ, or the eleven subtests. Substantial differences in the patterns of intellectual performance among the three groups were found when the adjusted means of the eleven subtests were compared to the actual mean subtest scores. These results suggest that the differences among ethnic groups on the WISC-III are more related to g, the general intelligence factor, rather than to narrower, more specific factors.

AN ANALYSIS OF THE PERFORMANCE OF A CLINICAL SAMPLE OF AFRICAN AMERICAN, CAUCASIAN, AND HISPANIC CHILDREN ON THE WISC-III

DISSERTATION

Presented to the Graduate Council of the

University of North Texas in Partial

Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Melissa L. Ewing, B.S., M.S.

Denton, Texas

December, 1998

ACKNOWLEDGMENT

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

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The History of Intelligence Testing

For centuries, philosophers and researchers have been intrigued by questions concerning differences among individuals, particularly intellectual differences. Investigation of the intellectual differences among individuals has focused specifically on differences in intelligence between the sexes and among racial, social, and ethnic groups. These investigations have led to a variety of interpretations which prompted test developers to create intelligence tests that most accurately and "fairly" assess intelligence in individuals, regardless of culture, ethnicity, or sex. One test developer, David Wechsler, created intelligence scales for adults and children that have received widespread acceptance and popularity among professionals and researchers. The most recently revised scale for children, the Wechsler Intelligence Scale for Children-Third Edition (WISC-III), retained the basic structure of the previous Wechsler Intelligence Scale for Children-Revised (WISC-R) and Wechsler Intelligence Scale for Children (WISC). The goals of this third revision included enhancement of the factor structure, improvement of subtests, and revision of norms. "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 the item-bias statistics for gender, ethnic, and regional bias" (Wechsler, 1991). The researchers reported that the very few items that were found to be biased were replaced. The purpose of the current study is to analyze the performance of a clinical sample of African American, Caucasian, and Hispanic children on the new WISC-III to determine if the test bias has been eliminated as claimed in the goals of the revision.

Early Theories of Intelligence and Intelligence Testing

In order to adequately investigate the newly revised test, it is necessary to review the history of intelligence testing and the theories, research, and debates surrounding the controversial topics of human intelligence and differences in IQ. It was not until the middle of the nineteenth century that researchers began to associate intelligence and individual differences in human beings.

Studying in England, Sir Frances Galton (1822-1911) developed his ideas about the nature of intelligence and individual differences in the context of Darwin's evolutionary theory (Hothersall, 1990). Regarded as the father of testing and measurement, Galton "originated two very important statistical concepts-regression to the mean and correlation-which permitted the psychometric field to flourish and develop" (Sattler, 1992). Galton's student, Karl Pearson, would later derive many other statistical analyses such as the product moment correlation coefficient, the multiple correlation coefficient, and the partial correlation coefficient (DuBois, 1970).

Galton's measurement of individual differences culminated in his establishment of the first psychometric clinic where individuals could have their mental and physical abilities

tested. Although his measurements were largely physical measures of visual and auditory reaction times, he was instrumental in creating the precedent for further psychological and intellectual testing. Galton also suggested that mental characteristics were distributed throughout the population with most people falling close to an average. He believed that extreme deviation from the average would become increasingly infrequent (Hothersall, 1990). As a result of these ideas, he advocated the improvement of the genetic stock of the human race through selective breeding. He suggested that this would eliminate lower intellectual ability and promote individuals with advanced intellectual ability.

At this same period in history, James M. Cattell (1869-1944), a former assistant of Galton, was developing a battery of instruments in the United States at the University of Pennsylvania. He had also studied under Wilhelm Wundt at the University of Leipzig and became interested in the study of individual differences (DuBois, 1970). His battery of instruments primarily assessed motor and sensory abilities. Although his measures were shown to have little predictive validity for most aspects of intellectual functioning, Cattell contributed to the field of test development by first using the term mental test and by suggesting that mental abilities should and could be experimentally evaluated (Sattler, 1992). Cattell's work helped to establish the foundations of mental measurement in the United States.

In addition, Cattell developed an innovative theory on the structure of intelligence. He proposed the concept of fluid and crystallized intelligence. Fluid intelligence he defined as essentially non-verbal, culturally-free mental ability. Crystallized intelligence referred to the acquisition of skills and knowledge that are strongly dependent on

exposure to culture and educational learning (Cattell, 1971). Cattell stressed that psychology must rest on the foundation of measurement and experimentation (Sattler, 1992). As a follower of Galton, Cattell also firmly supported the importance of genetic influences on individual differences and the encouragement of intermarriage among individuals of higher intellectual and mental abilities (Hothersall, 1990).

Charles Spearman (1863-1945), another researcher during this same time period, also developed an innovative theory which advocated a factor analytic study of intelligence. Basing his theories on actual work with the behavior of children, Spearman administered intelligence tests and evaluated the results using factor analysis. This technique was derived from the correlational work of Pearson. Spearman found that results on each individual task correlated with every other test in the battery in such a way that similarities between groups of scores were revealed (Spearman, 1904). He concluded that performance of an individual on any task depended first, on broad general intelligence (g) and secondly, on specific abilities peculiar to the task (Spearman, 1904).

In contradiction to the early theories of these individuals, some researchers suggested that intelligence could not be based on the physical measurement of sensory or motor abilities. A French researcher, Alfred Binet (1857-1911), proposed that the measurement of intelligence should focus on higher mental processes. He described the essential activities of intelligence as the ability "...to judge well, to comprehend well, to reason well..." (Binet & Simon, 1916). He criticized that the tests of Galton and Cattell were too heavily weighted in the direction of sensory function and failed to contain

sufficiently varied samples of measures pertaining to diverse mental faculties (DuBois, 1970).

In 1905, Binet developed what might be considered the first true test of what is now termed "intelligence." Together with his assistant, Theodore Simon, he created the Binet-Simon scale for the assessment of children. Binet intended that this test be used as a means of discriminating between children of inferior and retarded intelligence in the public schools of France. The test was administered to a innumerable group of French children, and in 1908, a revised edition of the test was published (Hothersall, 1990). Binet never used the terms mental age or intelligence quotient, although these terms were often associated with the Binet-Simon tests. Originally introduced by William Stern of Germany in 1911, the term mental age was later changed to intelligence quotient, IQ, by Lewis Terman. Binet and Simon were strongly opposed to the use of IQ and believed that this score would be dangerous and misleading. Despite this opposition, IQ became a standard way of describing the results of most intelligence scales.

The 1908 Binet-Simon was immediately successful in the world of intelligence measurement with twenty-two thousand copies distributed in three years (Hothersall, 1990). However, the test was often extended beyond the uses originally planned by Binet and Simon, in that it was often used as a mass test for a large group of individuals. Before the end of World War I, almost two million inductees into the United States Army were tested, and within two and a half years after its introduction in the United States, over four million children were also tested (Hothersall, 1990). The Binet-Simon scale was also

translated into the languages of other countries without adequate restandardization procedures using the individuals of these countries.

Henry H. Goddard (1866-1957), one of the men responsible for translating and introducing the Binet-Simon within the United States, did restandardize the test using individuals within the United States. However, Goddard also misused the 1908 Binet-Simon by conducting a controversial study of the Kallikak family (Hothersall, 1990). He found that this family was proof of the heritability of intelligence. Goddard found that one side of the Kallikak family contained descendants of ill-repute. These family members were characterized by feeblemindedness, criminality, promiscuity, and alcoholism. The other side of the family contained upstanding individuals such as doctors, lawyers, judges, and landowners. Using this study for support, Goddard advocated selective breeding to improve the stock of the human race.

Furthermore, Goddard also used intelligence tests to screen immigrants for the United States Government. With the influx of immigrants to the United States during the early part of the twentieth century, the government created laws that limited the numbers of immigrants (Hothersall, 1990). For example, no one could enter the U.S. if he was determined to be insane or feebleminded. Goddard originally claimed that he could determine feeblemindedness by simply looking at a person, but the government eventually hired him to administer intelligence tests to immigrants. Those who failed the test were determined to be feebleminded and sent back to their native country (DuBois, 1970).

Another researcher, Lewis M. Terman (1877-1956), investigated the strengths and weaknesses of the Binet-Simon test and concluded that a new revision was needed.

Terman's revision of the test included more extensive standardization procedures and added several new tasks to Binet's original set of tasks. He adopted the concept of the mental quotient from William Stern who had coined the term. However, Terman renamed the ratio the intelligence quotient (IQ) when he published the 1916 revision of the Binet-Simon scale (Terman, 1916). He attempted to evaluate and substantiate the validity of his revisions which sparked a fierce debate between himself and others such as Walter Lippmann, a well-known columnist and commentator in the early 1920s.

Lippmann expressed many concerns with the basis of intelligence tests even those created by Binet himself. He suggested that the Binet tests were not reliable because Binet seemed to arbitrarily choose the percentage of answers considered to be correct. A particular task was considered to be at or above a certain age level if a specific percentage of average children of that age could pass the task. Lippmann contended that this percentage was arbitrarily set by Binet (Block & Dworkin, 1976).

In addition, Lippmann stated that Binet randomly tried various tests and hoped that they tested intelligence. He argued that Terman, with his revision of the Binet-Simon test, used the same type guesswork. Lippmann claimed that each of Terman's mental age levels was an average performance of children to several arbitrary problems and puzzles (Block et al., 1976). He did not believe that these tests were truly measuring intelligence. Lippman was also particularly critical of a strictly genetic view of the nature of intelligence. He advocated that differences in early environment could make the comparison of different classes and races meaningless (Hothersall, 1990).

Definitions of Intelligence, Cultural Bias, and the Nature vs. Nurture Debate

It soon became apparent that the development of intelligence tests had 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, & Johnson, 1972).

Yet, 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.

This issue of cultural bias, however, cannot be addressed until one first examines the many definitions of intelligence. Some researchers have defined intelligence as 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). Binet 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 was abstract thinking (Samuda, 1975). Burt suggested that intelligence was innate, general, cognitive ability (Samuda, 1975), while David Wechsler believed that intelligence was 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, emphasized 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 believed intelligence to be the summation of learning experiences (Robb et al., 1972). His definition included all items measured by intelligence, aptitude, and achievement tests and allowed for the measurement of diverse abilities (Robb et al., 1972). Wesman suggested that it was important to consider not only innate abilities but all of the experiences during life that contribute to learning.

The definition by Wesman also incorporated the idea that "intelligent behavior is inescapably linked to and determined by the values and standards of society" (Samuda, 1975). This statement suggested 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, the intellectual differences between racially diverse individuals had been debated and discussed. Galton 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 rather than genetic differences might account for the findings (Edwards, 1971). As the controversy widened, Arthur Jensen (1969) published an article advocating that individual differences on intelligence tests had a high degree of genetic influence (almost 80 percent). In addition, he suggested that any social-class differences in IQ were also linked to a genetic component (Loehlin, 1975). He believed that the combined evidence suggested a reasonably sound hypothesis that genetic factors were strongly implicated in the intellectual differences between races (Jensen, 1969).

With Jensen's research, the arguments regarding the issue of test bias increased within the United States. Once again the debate focused on the influence of nature versus nurture. Both sides collected vast amounts of evidence to support their beliefs about the nature of intelligence. Some researchers used twin and sibling studies, adoption studies, and foster care studies to investigate the relation of heredity and intelligence (Samuda, 1975). Other researchers advocated that education and the social environment contributed 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 had the most influence on mental ability (Hothersall, 1990). Eysenck also believed that genetics accounted 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 were due to genetic differences

(Jensen, 1969). His studies found large mean difference between blacks and whites on tests measuring Spearman's general intelligence factor (Jensen, 1980). An eighteen point difference was found on the 1937 Stanford-Binet scale.

The environmentalists, on the other hand, refuted the arguments of hereditarians and suggested that intelligence was mainly influenced by environmental factors (Samuda, 1975). These individuals emphasized that unless children experience equal opportunities to learn what was required on intelligence tests, then differences in IQ scores would 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 differences (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 the testing of minority and deprived children has been that such standardized tests were 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 argued that specific item analysis revealed little bias within standardized tests

(Reynolds & Rechsly, 1983), while others showed evidence that the scores of minority children were significantly different from the scores of white children (Arinoldo, 1981; Munford, 1978; Munford and Munoz, 1980).

The term "bias" like the term "intelligence," however, has been difficult to define. One researcher, Anne T. Cleary (1968) stated 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 were labeled "unfair" if they were consistently lower for one group when compared to another group (Cleary, 1968).

Legislation Impacting the Use of Standardized Testing

Many organizations such as the Association of Black Psychologists (ABP) have advocated and supported the belief that standardized intelligence tests were 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.

Another similar case, Diana versus State Board of Education in 1970 (Sattler, 1992), investigated the use of standardized instruments with Mexican-American children.

A U.S. federal court ruled that testing procedures used in placing Mexican-American children in special education classes were invalid. This ruling stipulated that "linguistically different children would be tested both in their primary language and in English. . ." (Sattler, 1992). Furthermore, these children's cognitive abilities would be evaluated using ". . . primarily nonverbal tests and an interpreter would be used if a bilingual examiner was not available (Sattler, 1992)."

Both of these cases were particularly important in emphasizing the misuse of test data from standardized tests to place minority individuals in special classes. Many individuals felt that the test data were 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 argued that teachers did not consider environmental differences but only considered test score differences (Robb et al., 1972).

Within the educational system, it was not uncommon for students to be stratified into ability groups with students of similar intellectual achievements (Samuda, 1975). Because standardized intelligence tests were typically used for grouping students, minority children most often fell in lower groups as compared to white, upper-class students. This stratification system ranked children in terms of estimated potential (Samuda, 1975). This fact led to lower self-concept and perceived lowered expectations of teachers for children who were 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).

The ruling in 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 was 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 were primarily standardized on white middle-class groups of students, they produced inaccurate and misleading test scores when given to lower class and minority students (Hobson v. Hansen, 1967). As a result, the students were classified according to their socio-economic or racial status rather than their ability to learn (Hobson v. Hansen, 1967).

In 1975, Public Law 94-142 was passed which established that testing materials and procedures were to be selected and administered so as not to be racially or culturally discriminatory. PL 94-142 focused on the education of handicapped children but was also intended to apply to all children including children with physical impairments and learning disabilities. This legislation also stated 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 was such that education in regular classes with the use of supplementary aids and services could not be achieved satisfactorily (PL 94-142, 1975). This legislation hoped to increase the equality of educational opportunities for all children and to reduce the segregation of such children based on standardized test materials.

More recently, the 1980 Parents in Action on Special Education, PASE, case came to opposite conclusions of the Larry P. case (Kamphaus, 1993). This case in Chicago investigated primarily the bias against Latino children placed in educable mentally retarded, EMR, classrooms. In this instance, the judge determined that ". . .the bias identified in the Wechsler and Binet items was not substantial" (Kamphaus, 1993). Furthermore, another case, Marshall v. Georgia (1984), upheld the instructional grouping and placement practices in the rural school districts of Georgia.

Development of the Wechsler Scales

Research on Test Development and Bias through the WISC-R

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 separated natural intelligence from that contaminated and obscured by education (Cattell & Cattell, 1959).

Another test was the Progressive Matrices Test which was developed in Great Britain by Raven (1965). This nonverbal test consisted of sixty matrices or designs from each of which a part had been removed (Raven, 1965). The Leiter International Performance Scale was another nonverbal test in which both the examiner and examinee performed rather than verbalized 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 seemed to measure only a very limited aspect of intelligence and were not "culturally fair" to more than a few isolated cultures (Robb et al., 1972). The tests often failed to give much useful information about the individual being tested and often did 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; however, they have not succeeded in replacing the traditional standardized intelligence tests. Test developers continued to be concerned with issues of test bias and culturally fair intelligence tests. One such test developer was 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 (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, 1992). 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, 1992).

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, 1992). Within the standardization sample, the differences between mean IQs for boys and girls was less than three points. Therefore, these differences were determined to be too small to reveal any clinically significant difference in boys and girls. Most research on gender differences has revealed consistent but small differences between males and females. Kaufman (1979) reported that girls scored one and one-half points higher than boys on the Coding subtest of the WISC-R. Boys scored higher by about half a scale score point on all of the other Performance subtests (Kamphaus, 1993). Comparisons of these differences revealed almost equal Performance IQs for boys and girls.

Mean IQ scores for white children, however, were found to be one standard deviation higher than black children (Sattler, 1992). Furthermore, children whose parents were professional or technical workers obtained higher IQs than children whose parents

were unskilled workers. Differences in IQ for children in urban versus rural areas were small. Regional differences were found in the South and West with children in the South having IQ scores lower than children in the West (Sattler, 1988).

A large body of research is available on the WISC-R (for summary, see Appendix, Table 1). In particular, a large amount of research has been dedicated to the study of the factor structure of the WISC-R and to the investigation of the differences found between different ethnic groups on the WISC-R. Research on the factor structure of the WISC-R is particularly controversial surrounding the issue of differences between ethnic groups.

David C. Geary and Randolph H. Whitworth (1988) investigated the potential differences in factor structure of the WISC-R for Anglo- and Mexican-American children. These authors contended that if the factor structure is different for minority and white children, then ". . . the subtests of the WISC-R might not assess the same underlying cognitive abilities in these . . . groups" (Geary et al., 1988). The "Freedom from Distractibility" factor is of particular concern among ethnically diverse groups. These researchers found that the factor structure of the WISC-R was nearly identical for the two groups. The same subtests appeared to contribute to the same factors in both groups. Therefore, they concluded that "clinical and educational interpretations of the cognitive processes that underlie performance on the WISC-R subtests probably should be the same for Anglo- and Mexican-American children" (Geary et al., 1988). Some differences were found on language-related measures. These differences were attributed to bilingual and bicultural environments of the Mexican-American children. Therefore, the subtests

defining the "Verbal Comprehension" factor should be interpreted with caution for Mexican-American children.

Other factor analytic studies of the WISC-R with African American children, Mexican-American children, and lower-middle-class children also indicated that the factors were generally defined by the same subtests as in the standardization sample (Carlson, Reynolds, & Gutkin, 1983; Dean, 1980; Johnston & Bolen, 1984). Oakland and Feigenbaum (1979) found a high correlation for Anglo- and Mexican-American children on the WISC-R factors. They found a correlation of .97 for these two groups on "g" factor loadings. Finally, there have been consistent findings that the "Freedom from Distractibility" factor accounts for only a small amount of the variance with the largest portion of variance accounted for by the Verbal and Performance Scales (Kamphaus, 1993).

In addition to factor analytic studies, researchers focused on the differences between the scores of a variety of ethnic groups on the WISC-R. Reynolds and Rechsly (1983) supported the claims of Wechsler that his tests contained 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 was assessed on the internal criteria of reliability and a number of order-of-item-difficulty measures. Sandoval (1979) found that minority children did not differ from 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.

Furthermore, Vance, Hankins, and McGee (1979) found evidence of unexpected differences between minority children and white children. They found that minority children performed better on verbal tasks than white children. These researchers hypothesized that unexpected findings might be attributable to homogeneity of their sample. Although they suggested that their findings be interpreted cautiously, the results indicated that the WISC-R seemed to be unbiased and that some minority children actually performed better than white children.

Studies of predictive bias on the WISC-R have also supported the use of intelligence scales with minority and non-minority children. Several researchers have found that the WISC-R was able to predict school achievement equally well for both Caucasian and minority children. Rechsly and Rechsly (1979) found that the WISC-R was an adequate predictor of school achievement based on the Metropolitan Achievement Test. They stated that "the WISC-R Full Scale IQ appears to have very much the same meaning for Mexican-American and Anglo students, despite differences in the mean IQs" (Rechsly et al., 1979). These researchers also investigated the three factor scores for the WISC-R and found that Verbal Comprehension was the best predictor of reading and math achievement.

Oakland (1980) found similar results when comparing Anglo, Hispanic, and African American children. They found that the three scales of the WISC-R were equally correlated with the Iowa Test of Basic Skills for all three ethnic groups. They correlated

these three scales with reading and math achievement scores and concluded that the WISC-R was an unbiased predictor of school achievement (Oakland, 1980).

Similarly, Dean (1979) concluded that the WISC-R was an unbiased predictor of school achievement. This researcher found that predictive validity coefficients for verbal subtests were generally higher than for performance subtests. Dean (1979) also used the Iowa Test of Basic Skills as the criterion measure for school achievement.

In contrast to this research, Reynolds and Gutkin (1981) found significant differences in their study of 285 matched pairs of African American and Caucasian children. Their analysis of the results on the WISC-R revealed that Caucasian children exceeded African American children 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 African American 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. They also stated that the differences in Verbal IQ "reflects the extent of acculturation, while the Performance IQ reflects actual ability" (Munford et al., 1980).

Arinoldo (1981) found that African American and Caucasian children differed significantly in their Wechsler Full Scale IQs. He found these differences in preschool and school-age children. Taylor and Richards (1991) found differences among African American, Hispanic, and Caucasian children when they looked at patterns of intellectual performance. Since factor analyses of the WISC-R have found similar factor structure

across ethnic groups, these researchers concluded that the WISC-R was measuring similar constructs for children of different ethnic status. Therefore, they researched the intellectual patterns of ability within these three ethnic groups by controlling the Full Scale IQ (Taylor et al., 1991).

These researchers found that Hispanic children seemed to perform better on visual-spatial tasks, Caucasian children performed better in abstract thinking and knowledge of facts, and African American children performed better on verbal tasks (Taylor et al., 1991). These children also varied on the three factors with African American children highest on Verbal Comprehension, and Hispanic children highest on Perceptual Organization and Freedom from Distractibility. They concluded that ". . . children with the same Full Scale IQ do not necessarily have the same pattern of cognitive strengths and weaknesses" (Taylor et al., 1991).

In another study, Munford, Meyerowitz, and Munford (1980) compared the results of African American and Caucasian children on the WISC and WISC-R. They found no differences in the scores of Caucasian children on the WISC and WISC-R. However, they found that the African American children scored significantly lower on the WISC-R than on the WISC on all three IQ measures. They concluded that their results indicated 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 African American 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 produced lower IQs for certain minority children may also support the hypothesis that these low IQ scores would predict poor school performance (Cleary, Humphreys, Kendrick & Wesman, 1975). Therefore, more African American children than Caucasian 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 African American 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 African American culture. Researchers showed that this test had questionable validity and had low correlation with a currently accepted achievement test. This 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 African American 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 black children did not appear to make these tests more culturally fair. Quay (1971) translated the Stanford-Binet into "Black English" and found that African American 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 did suggest that some tests may possess a bias that cannot be eliminated.

Development and Research on the New WISC-III

Due to these 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. In 1991, Wechsler published his latest version of his intelligence scales for children, the Wechsler Intelligence Scale for Children--Third Edition. This new scale was revised seventeen years after the publication of the successful WISC-R. Researchers stated that some of the goals of the WISC-III were to periodically update norms, to maintain the basic structure and content of the WISC-R, to enhance the factor structure underlying the WISC-R, and to improve various subtests (Wechsler, 1991). Research indicated that the very few items that were found to be biased were replaced in the WISC-III (Wechsler, 1991). Seventy-three percent of the WISC-R items were retained in the WISC-III. Although no data was available within the manual, the manual stated that item-bias analyses were conducted with the WISC-III

"tryout data" (Wechsler, 1991). Test developers also 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.

The WISC-III contains thirteen subtests, including Information, Similarities, Arithmetic, Vocabulary, Comprehension, Picture Completion, Coding, Picture Arrangement, Block Design, Object Assembly, Digit Span, Symbol Search, and Mazes. The last three of these subtests are considered supplementary, while the other ten subtests are used to formulate the Full Scale IQ. Kamphaus and Platt (1992) computed the new subtest specificities for the WISC-III using a method previously described in 1979. This information provided by these authors fills ". . . the interpretive void by providing a convenient table of subtest specificities for the WISC-III" (Kamphaus et al., 1992). This information provided examiners with the knowledge of how accurately the interpretation of individual subtest scores measured unique abilities.

Kamphaus et al. (1992) found that the subtest specificities for all age groups of the WISC-III were similar to those specificities reported for the WISC-R. Object Assembly had the lowest specificity, suggesting that this subtest cannot be interpreted as a single strength or weakness of a particular child. Digit Span, on the other hand, had the highest specificity and seemed to represent a unique ability in and of itself (Kamphaus et al., 1992).

Kramer (1993) also 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. One table provided the range of scores that can be used for interpretation of a specific obtained subtest score. This range represented the range of true scores or 1 SEM below and above the estimated true score. Another table provided a classification system of verbal descriptions for the individual subtest scores. Kramer (1993) concluded that his study suggested a more conservative approach to interpretation of subtest profiles.

The standardization procedure for the WISC-III included four geographical regions, both sexes, several ethnic/racial groups, and the entire socioeconomic status (SES) range. Furthermore, researchers tested an oversample of four hundred minority children in addition to the standardization sample and analyzed their scores for bias (Wechsler, 1991). This sampling procedure was notably improved from the WISC-R, which stratified race only according to "White versus Non-white" (Wechsler, 1974).

The three WISC-III IQ scales have outstanding reliability with internal consistency reliability coefficients of .89 or greater for the entire age range covered in the standardization group (Sattler, 1992). However, the internal consistency reliability coefficients were lower for the subtests, ranging from .69 to .87. Wechsler (1991) reported the average standard error of measurement (SEM) for the WISC-III Full Scale IQ, Verbal IQ, and Performance IQ as 3.20, 3.53, and 4.54, respectively. The SEM for the subtests in scaled score points ranged from 1.08 to 1.45 (Sattler, 1992). The SEM

provided 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 three hundred and fifty-three children who were tested twice (Wechsler, 1991). The WISC-III possessed 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). However, Performance IQ increased by 11.5 to 13.5 points. "The large retest gains on the Performance Scale are of major concern when you readminister the WISC-III to children after a period of 2 to 9 weeks. For longer periods of time, however, gains on retest are likely to be lower because practice effects tend to diminish over time" (Sattler, 1992).

Slate, Jones, and Saarnio (1997) confirmed the stability of the WISC-III IQ scores with their study of special education students. They discovered nonsignificant differences between the scores of the students on two administrations of the WISC-III three years apart. The children's scores were obtained during initial evaluation and routine 3-year re-evaluations (Slate et al., 1997). These authors concluded that test-retest scores were significantly correlated for their sample of special education students.

Wechsler (1991) reported that the WISC-R and WISC-III were also highly correlated with .90 and .89 for Verbal IQ and Full Scale IQ, respectively. The correlation between Performance IQ was slightly lower ($r = .81$). A comparison of the mean Full Scale IQ score showed that the WISC-III FSIQ was approximately five points less than the WISC-R FSIQ. The WISC-III VIQ and PIQ scores were two and seven points less

than corresponding WISC-R scores. In addition, Wechsler (1991) found similar results with a clinical sample of one hundred and four children. The correlations were .86, .73, and .86 for Verbal, Performance, and Full Scale IQs, respectively. Full Scale IQs were 5.9 points lower on the WISC-III than the WISC-R in this clinical sample (Wechsler, 1991).

Specific changes in the WISC-III from the WISC-R included a new supplementary subtest, Symbol Search. A new factor called Processing Speed was added because the Freedom from Distractibility factor emerged as a weak factor in the WISC-R (Sattler, 1992). Modifications in the scoring and administration of the WISC-III were changed as well as the addition of full color illustrations within certain subtests. Bonus points were now awarded for speed on the WISC-III, and the number of items was increased on many of the subtests (Sattler, 1992).

As with any new assessment tool, authors have evaluated and critiqued the positive and negative aspects of the newly revised WISC-III (for summary, see Appendix, Table 1). Alan S. Kaufman (1993) summarized and commented on the changes found within the WISC-III. He commended the new artwork and addition of color within the test, applauding the efforts made by the publisher to avoid biased items. Kaufman (1993) pointed out that extensive and rigorous means were employed to evaluate item bias. However, he questioned the elimination of some of the items with clinical content, describing this action as one that "David Wechsler never would have allowed . . . " (Kaufman, 1993). Kaufman (1993) indicated that many individual examiners have complained that "favorite questions" from previous Wechsler scales are no longer present in the new revision. Furthermore, Kaufman indicated that the WISC-III seemed to place a

substantial emphasis on a person's problem-solving speed. He stated that this emphasis may unintentionally penalize children with reflective cognitive styles or learning disabilities. Even the new supplementary subtest and new fourth factor were based on processing speed. Kaufman (1993) concluded that the WISC-III was a superior instrument with "comprehensive evidence of reliability and validity." However, he believed that interpretation may be hindered by the publisher's decisions to eliminate items with clinical content and to increase emphasis placed on processing speed. Finally, Kaufman's main disappointment was that a special subtest was not devised to provide a "fairer evaluation of the intelligence of children with school problems and of children from different cultural or linguistic backgrounds" (Kaufman, 1993).

Thelma Blumberg (1995) and Post and Mitchell (1993) also assessed the attributes of the WISC-III, concluding that the benefits of the WISC-III outweighed its flaws. Blumberg (1995) emphasized the use of the WISC-III with special populations, citing its excellence as a diagnostic tool for mental retardation and learning disabilities. Post and Mitchell (1993) emphasized the decline in IQ scores from the WISC-R to the WISC-III. They commented on the impact of revised intelligence measures on special education services.

Of particular concern was the fact that declines in IQ scores could result in decreased numbers of learning disabled children who qualify for services and increased numbers of children who qualify for special education services. These facts suggested that conclusions for special education services should not be based solely on individual test scores. Post et al., (1993) stated that the fluctuations in numbers of children qualifying for

special educational services merely indicated that the WISC-R overestimated the ability of these children. The authors concluded that these factors will need to be considered when the WISC-III is utilized by practitioners; however, they conclude that the field will ultimately benefit from this new intellectual assessment tool (Post et al., 1993).

Finally, Steve Edelman (1996) reviewed the changes found in the WISC-III and concluded that the WISC-III was not entirely a success. "Although there are substantial improvements in presentation and scoring, the test leaves much to be desired in the area of presenting a new factor, and the increased importance of speed for adolescents to score high on certain subtests" (Edelman, 1996). In agreement with Post et al., (1993), this author pointed out the possible negative impact of decreased IQ scores between the WISC-R and WISC-III for placement and re-evaluation of exceptional children. He stated that the impact would be greatest in school systems that rely on ability test score/achievement test score discrepancies (Edelman, 1996).

In addition to these authors who have critiqued the overall changes within the WISC-III, Husband and Hayden (1996) specifically studied the impact of the addition of color within the WISC-III. They found that color did not negatively or positively impact the subtest, Picture Completion. However, the authors did find a distinct preference for colored items as stated by the children that they evaluated. No expected main effect differences were found between regular or remedial students. Husband et al., (1996) concluded that the "students' obvious preference for color . . . seems to justify the addition of color to these, and potentially other, standardized tests."

Literature on the revised WISC-III has also included investigations of the newly defined four-factor structure. This four factor model included Verbal Comprehension (sum of scaled scores on Information, Similarities, Vocabulary, and Comprehension), Perceptual Organization (sum of scaled scores on Picture Completion, Block Design, Object Assembly, and Picture Arrangement), Freedom from Distractibility (sum of scaled scores on Arithmetic and Digit Span), and Processing Speed (sum of scaled scores on Coding and Symbol Search). Sattler (1992) stated that the four factor model needed more empirical evidence, and he preferred to use a three factor model that excluded the factor, Freedom from Distractibility. As with the previous versions of Wechsler's scales, the factor structure of the WISC-III has been an area for debate.

For example, research on the WISC-III reveals discrepancies over the Freedom from Distractibility factor. Wechsler (1991) reported the evidence to suggest four factors for the WISC-III. Roid, Prifitera, and Weiss (1993) also confirmed the four factor structure of the WISC-III in their replication of the study of the factor structure found in the manual of the WISC-III. Hishinuma and Yamakawa (1993) studied the factor structure as applied to exceptional students and those "at risk". They found that the four-factor model was supported; however, Picture Arrangement fell within the domain of the Processing Speed factor, contrary to prior results. Hishinuma et al., (1993) stated that Processing Speed appeared to be associated with visual-motor/sequential processing within a timed component. Furthermore, since the Freedom from Distractibility factor now contained only Arithmetic and Digit Span, this factor may now be more representative of a math-achievement component.

Other authors have also examined the factor structure of the new WISC-III and questioned the strength of the third and fourth factors, Freedom from Distractibility and Processing Speed. Kamphaus, Benson, Hutchinson, and Platt (1994) studied three different models as applied to the WISC-III--Wechsler's original two-factor conceptualization, Kaufman's three-factor model, and the four-factor model proposed in the WISC-III manual. They stated that "although the verbal comprehension and perceptual organization factors of the WISC-III remain intact, the freedom from distractibility and processing speed factors by contrast are much weaker due to the Mazes, Digit Span and Coding subtests" (Kamphaus et al., 1994). These authors emphasized the need for more empirical data to support the two questionable factors, warning that practitioners should use these factors cautiously when making interpretations about exceptional children.

Reynolds and Ford (1994) also investigated the factor structure of the WISC-III, focusing on the three-factor structure across the eleven age levels between 6 1/2 and 16 1/2 years. As in studies of the WISC-R, these results indicated that a three-factor structure was "stable across statistical techniques and across the age range of the instrument when Symbol Search is deleted" (Reynolds et al., 1994). These authors stated that the fourth factor may be appropriate but lacked empirical support. They pointed out that the traditional three-factor solution may be preferable.

Slate and Jones (1995) more recently investigated the factor structure of the WISC-III as applied to African American students undergoing special education re-evaluation. They found that the factor structure was generally supported for this

population with unrotated factors from the analyses supporting the Full Scale, Verbal, and Performance IQs. Generalizations from this study should be made with caution because of the small sample size and variety of diagnoses among the individuals included in the study.

Additional research has not only included factor analytic studies but also investigated the validity of the WISC-III, comparing the new test with the WISC-R and other intellectual and achievement measures. 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 extended beyond IQ to the level of the item.

Other authors have also compared the newly revised WISC-III with different intelligence tests, including the Stanford-Binet IV (SB:IV). Prewett and Matavich (1994) administered these two tests to a group of lower SES inner city children referred for psychoeducational evaluations. While they found a high correlation of .81 between the Full Scale IQ of the WISC-III and the SB:IV Test Composite, they discovered that the two tests did not necessarily give similar diagnostic impressions. Prewett et al., (1994) found that the WISC-III Full Scale IQ was significantly lower than the SB:IV Test Composite by 9.4 points. Furthermore, the authors found that the two tests often gave different profiles of nonverbal and verbal performance (Prewett et al., 1994). They concluded that "an individual's diagnosis and special education placement often will differ depending on the test given" (Prewett et al., 1994).

Lukens and Hurrell (1996) also compared these two intelligence tests, WISC-III and SB:IV, with a group of mildly retarded children. The comparison revealed high and significant correlations between the two tests; however, the Full Scale IQ of the WISC-III was significantly lower than the SB:IV Composite Score for the majority of the individuals in this study. In agreement with Prewett et al., (1994), Lukens et al., (1996) concluded that multiple criteria are needed to determine the qualification and educational placement of children with disabilities.

Finally, Lavin (1996) and Rust and Lindstrom (1996) evaluated the WISC-III and SB:IV with nonreferred student populations. They found similar high and significant correlations between the two tests of .81 for the WISC-III Full Scale IQ and the Test Composite of the SB:IV. However, these two studies did not find significant differences between the mean scores. Both studies cautioned that the small sample sizes may have contributed to such results, but the significant correlations between the two tests still supported earlier research (Lavin, 1996) and (Rust et al., 1996).

In a similar manner, Caravajal, Hayes, Miller, Wiebe, and Weaver (1993) studied differences between the WISC-III IQ scores and the Peabody Picture Vocabulary Test-Revised (PPVT-R) for third through fifth graders. Results revealed moderate correlations between the two tests, and the authors concluded that these correlations were "sufficient to warrant the use of the PPVT-R as a brief screening test for children in Grades 3, 4, and 5" (Caravajal et al., 1993).

Slate, Jones, Graham, and Bower (1994) and Slate (1995) both investigated the criterion validity of the WISC-III and the PPVT-R for children with specific disabilities.

In both studies, the research indicated moderate and significant correlations between the two measures. Therefore, these two measures assess similar constructs, providing evidence of criterion validity of the WISC-III and the PPVT-R for mentally retarded or learning disabled children (Slate, 1995).

Significant correlations were also found between the WISC-III and the Kaufman Assessment Battery for Children (K-ABC) (Rust & Yates, 1997). These authors reported a correlation of .61 between the WISC-III and K-ABC. While they found no significant difference between the mean scores of the two measures, they did suggest that the variability among the scores of their participants indicated that in some cases differences between the tests can be great. For example, one participant had a Full Scale IQ score 37 points higher than the Mental Processing Composite of the K-ABC (Rust et al., 1997).

Recent researchers have indicated that the WISC-III is highly correlated with other intelligence measures; furthermore, researchers have also attempted to discover the relationship between the new WISC-III and the previous WISC-R. 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 mean 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 concluded 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).

Dumont and Faro (1993) also found decreases in IQ scores from the WISC-R to the WISC-III for 41 learning disabled children. They found mean score differences of 7.7, 6.7, and 8.2 for Full Scale IQ, Verbal IQ and Performance IQ, respectively. These mean score differences varied from the differences of 4, 2, and 7 points, respectively, as reported in the WISC-III manual. Hager (1992) also analyzed differences between the WISC-R and WISC-III mean scores for special education students and found decreases on all three IQ scales. When the data were analyzed by race, mean score changes for African American children were significantly greater on all three IQ scales. As with Dumont et al. (1993), the decreases in IQ for all three scales were larger than reported in the WISC-III manual.

Nagle and Daley (1994) 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 emphasized the need for further research, they suggested 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).

Several other researchers also found similar results, revealing high correlations between the two Wechsler tests and declines in IQ scores between the WISC-R and the WISC-III (Bolen, Aichinger, Hall, & Webster, 1995; Gunter, Sapp, & Green, 1995; Graf & Hinton, 1994; Lyon, 1995; Slate, 1995; Slate & Saarnio, 1995;). All researchers examined groups of students with learning disabilities or mental retardation. Most found

results similar to the comparisons discussed within the WISC-III manual; however, Lyon (1995) and Slate (1995) found slightly greater declines in IQ scores than those reported in the manual. Although most of these researchers used small samples of predominately Caucasian children, Gunter et al. (1995) studied a small group of African American urban learning disabled children and found similar results.

Two of these studies, in particular, discovered additional differences within the comparisons of the WISC-R and WISC-III. Slate et al. (1995) found that children with specific learning disabilities (SLD), mental retardation (MR), and no disabled classification (NC) differed in their increases and decreases from the WISC-R to the WISC-III. These authors pointed out that not all children decrease in scores. Some children's scores actually increased. In this study, a greater percentage of the SLD group increased their IQs, while a greater percentage of the MR group decreased their IQ scores. The NC group had few children that either increased or decreased their scores from one test to the other (Slate et al., 1995).

In a similar study, Slate (1995) investigated only the WISC-III scores of three groups of students--SLD, MR and NC. He found that, as indicated in the WISC-III manual, all three groups showed discrepant IQ and Index scores. However, the magnitude of the discrepancies was smaller than reported by Wechsler. In addition, all three groups in the study demonstrated higher Performance IQs than Verbal IQs (Slate, 1995). He concluded that these discrepancies, however, were too small to serve as adequate diagnostic indicators.

Another unique result was discovered by Graf et al. (1994). In their study of eighty-four regular and special education students, only Full Scale IQ and Performance IQ decreased significantly from the WISC-R to the WISC-III. They concluded that the WISC-III appeared to be more difficult and seemed to cluster individuals at the lower IQ ranges. These researchers cautioned that this trend could more frequently place students in classes for the educable mentally retarded (Graf et al., 1994).

Finally, Ewing (1994) found significant differences in the performance of a clinical sample of children on the WISC-R and WISC-III. In particular, this researcher found that the scores of African American males were lower on the WISC-III in comparison to the scores of Caucasian children. The scores of the Caucasian males revealed no differences between the WISC-R and the WISC-III. In fact, this group's scores showed slight increases between the two tests, while the scores of the African American males showed the characteristic decreases between the two tests. This study concluded that further research was needed to determine what factors were contributing to differences between minority and Caucasian children.

All of these researchers agree with the warnings stated by Post and Mitchell (1993). They speculated that decreases in scores on the WISC-III for special education students may impact their placement in special classes. All of these authors emphasized that, although declines in scores from the WISC-R to the WISC-III are typical of new revisions of intelligence tests, these declines may be detrimental to a child's placement in special education classes. Children with learning disabilities may no longer qualify for special classes due to these decreases in scores. Their lower IQ scores may no longer be

significantly discrepant from achievement scores. However, the number of children classified as mentally handicapped may increase due to the decrease in IQ scores from the WISC-R to WISC-III (Post et al., 1993). All recent researchers caution that these problems be considered when children are being assessed for special education programs. They also emphasized the need for continued research on the newly revised WISC-III with larger samples of ethnically diverse children.

CHAPTER 2

METHOD

Comparison of WISC-III Scores for Minority and Caucasian Children

Introduction

It appears that intelligence test development may never completely eliminate the impact of cultural and educational 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 sample." In other words, it may be important to study the Wechsler Intelligence Scale for Children-III and determine whether improvements in the WISC-III have successfully reduced the culturally-loaded materials. It may be that several factors such as cultural biases, environmental factors, and genetic differences interact to reflect the observed differences between various ethnic and disability groups.

Hypotheses

In the present study, various aspects of the WISC-III were investigated and analyzed. This study focused on comparisons between Caucasian, African American, and Hispanic children, males and females. The goal was to determine if test developers successfully reduced some of the culturally loaded items that they claimed to have found and eliminated in the WISC-III (Wechsler, 1991). Several methods were used to

investigate differences between the ethnic groups. First, analyses were used to determine if any differences in gender, age, ethnicity, parent education, family income or primary diagnosis existed among the children for the three IQ measures and the eleven subtests of the WISC-III. Elimination of biased items on the WISC-III would suggest that no differences would be found among the children on any of the scores for any variable such as gender, age, ethnicity, parent education, family income, or primary diagnosis.

Hypothesis 1

There would be no gender or age differences among the individuals on the WISC-III subtests and IQ scores.

Hypothesis 2

There would be no differences among Hispanic, African American or Caucasian children on the WISC-III subtests and IQ scores.

Hypothesis 3

There would be no differences on the WISC-III subtests and IQ scores for children with different family incomes.

Hypothesis 4

There would be no differences on the WISC-III subtests and IQ scores for children whose parents have different educational levels.

Hypothesis 5

There would be no differences on the WISC-III subtests and IQ scores for children with different primary diagnoses.

Secondly, this author investigated the patterns of intellectual abilities between the ethnic groups when IQ is held constant as the covariate. Differences between the ethnic groups on the ten subtests were investigated based on the patterns of intellectual performance. If the WISC-III improvements have reduced bias for ethnic minorities, then no differences should be found in the patterns of intellectual abilities for the three ethnic groups when IQ is held constant.

Hypothesis 6

No differences in the patterns of intellectual abilities would be found for the three ethnic groups, Hispanic, African American, and Caucasian, when Full Scale IQ is held constant.

Subjects

Demographic information was obtained from the psychological folders of 241 children, males and females, between the ages of seven and sixteen (see Appendix, Table 2). This study included 120 females (51 Caucasian, 35 African American, and 34 Hispanic) and 121 males (51 Caucasian, 35 African American, 35 Hispanic). Gender, ethnicity, age, socioeconomic status, parent education, and diagnoses were recorded for each individual. Socioeconomic status was recorded as combined family income, and parent education was divided into categories as specified in the WISC-III manual (Wechsler, 1991). Categories included completion of 8th grade or below, 9th - 11th grade, 12th grade, 13-15 years, or 16+ years. Education was recorded for both parents when available.

All of the subjects were clients at a local metropolitan children's clinic. These youth were receiving psychological services primarily for Attention-Deficit Hyperactivity

Disorder, Oppositional Defiant Disorder, or specific developmental disorders such as Developmental Reading and Developmental Arithmetic Disorders. Every child tested at this metropolitan clinic, however, was not included within this study. After a specified number of Caucasian children were found, research efforts focused on finding minority children to fill the remaining cells with an adequate and fairly equal number of children to provide valid and intrepretive results. In addition, children with Pervasive Developmental Disorders, Autism, Psychotic Disorders, and Major Clinical Depression were excluded from this study.

All of the subjects were administered the WISC-III between January, 1992 and September, 1995. Due to the fact that the WISC- III included 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 included Picture Completion, Information, Coding, Similarities, Picture Arrangement, Arithmetic, Block Design, Vocabulary, Object Assembly, and Comprehension. For the purposes of this study, scores from the subtest Digit Span were also analyzed, although this test was not used to derive the Full Scale IQ score. All tests were administered by a Licensed Psychologist, Psychological Associate, or Psychology Intern. Both the Psychological Associate and Intern were supervised by the Licensed Psychologist. Furthermore, the test scores were also reviewed by the primary author to eliminate the possibility of error.

Analyses

Means and standard deviations were computed separately for the three groups

(Caucasian, Hispanic 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, and the eleven individual subtests. For the WISC-III, the three IQ scales have a mean of 100 and a standard deviation of 15.

A t-test of independent samples was used to determine if there were any significant gender differences between the overall sample on IQ scores and individual subtest scores. T-tests for independent samples were also used to determine any gender differences on the IQ and subtests scores within the three ethnic groups. The data was subjected to separate analyses of variance to determine any significant differences between the different ethnic groups based on family income, father's educational level, mother's educational level, or primary diagnosis.

A multivariate analysis of variance was used to determine any significant differences between the ethnic groups on the eleven subtests, Verbal IQ score, Performance IQ score, or Full Scale IQ score. Univariate ANOVAs were conducted as follow-ups for significant MANOVAs. Further evaluations included a multivariate analysis of covariance investigating the difference among the three ethnic groups on the WISC-III IQ scores and subtest scores when family income was designated as the covariate.

An additional multivariate analysis of covariance was performed controlling for the possible effects of parental education level. A nonparametric correlation was conducted as follow-up on the relationship between mother's educational level and the three overall IQ scores. A multivariate analysis of covariance was also employed to investigate the

differences among the ethnic groups when both mom's education and income are held constant as the covariates.

Finally, a multivariate analysis of covariance investigated the patterns of intellectual performance among the groups. The subtest scores were used as the dependent measures with the FSIQ held constant as the covariate. Differences in VIQ and PIQ were also investigated by using a multivariate analysis of covariance controlling for the effect of the FSIQ. This analysis was modeled after the research by Taylor and Richards (1991).

CHAPTER 3

RESULTS

A comparison of the demographic variables among the three racial groups produced a mixture of results. No significant differences were found when comparing the mean ages of the three groups, while a significant difference ($p < .01$) was evident between the groups on the basis of income with a mean income for the African American children of $10,733.00, the Caucasian children of $24,971.00, and the Hispanic children of $17,723.00 (see Appendix, Table 2).

There was also no evidence of overall gender differences in IQ when comparing the sample as a whole (see Appendix, Table 3). The sample as whole was not significantly different on any of the IQ scores or subtests except coding and object assembly. The females performed significantly better on coding ($p < .01$) than the males, while the males performed significantly better than the females on object assembly ($p < .05$).

Similarly, within each individual racial group, there were no significant differences between males and females on the Verbal, Performance, and IQ scales. For the African American children, there were also no significant differences between the males and females on any of the eleven subtest scores. The Caucasian males and females were significantly different only on the picture completion subtest ($p < .01$) with the Hispanic males and females only significantly different on the coding subtest ($p < .01$).

A preliminary analysis of variance comparing the three ethnic groups on the overall IQ scores and eleven subtests scores indicated a statistically significant difference ($p < .01$) among the groups on all overall IQ scores (see Appendix, Table 4) and all subtests excluding digit span, coding, and arithmetic (see Appendix, Table 5). The difference among the groups on arithmetic was significant at the ($p < .05$) level with no significant differences among the groups on digit span and coding.

Furthermore, an analysis of variance comparing the three ethnic groups split by gender also found significant difference on all three IQ measures for the females (see Appendix, Table 6); however, the males were significantly different only on the Verbal and Full Scale IQ scores. Investigation of the subtest scores for each gender produced significant differences for the African American, Caucasian, and Hispanic males on all subtests excluding arithmetic, coding, and digit span (see Appendix, Table 6). The three groups of females were significantly different on all subtests except arithmetic, coding, digit span, and picture arrangement.

In order to further delineate specific factors that might contribute to the differences between the three ethnic groups, an analysis of variance was used to determine the interaction of various demographic variables on the IQ scores of these three groups. An analysis of variance investigated the impact of father's education on Verbal, Performance, and Full Scale IQ. No significant differences were found in Verbal, Performance, or Full Scale IQ for African American males or Hispanic males when the educational level of their fathers was considered. Only Caucasian males showed significant differences in their Performance and Full Scale IQ based on their father's level

of educational attainment (see Appendix, Table 7). Higher educational achievement seemed to suggest a higher IQ score, particularly with Full Scale IQ. There seemed to be no correlation between father's education and IQ for any of the female groups (see Appendix, Table 8).

The impact of mother's educational level was also evaluated. There were significant differences on IQ measures for all three racial groups (see Appendix, Table 9). However, when divided by gender, there were fewer significant differences based on mother's education (see Appendix, Tables 10 & 11). African American males showed significant differences on Verbal and Full Scale IQ, while Caucasian males showed significant differences on all three IQ scores. Hispanic males and all three of the female groups demonstrated no significant differences on any of the IQ measures.

A follow-up correlation using Spearman's rho was performed to determine the relationship between mother's education and the three IQ measures. An overall significant correlation was found between mother's educational level and Verbal IQ scores ($r = .382$; $p < .01$), Performance IQ scores ($r = .303$; $p < .01$), and Full Scale IQ scores ($r = .369$; $p < .01$). As found with the ANOVA, a significant correlation existed between mother's education and African American males VIQ score ($r = .376$; $p < .05$) as well as all three IQ scores for Caucasian males ($p < .01$).

In order to factor out the impact of mother's education on the overall difference between the groups, the data were subjected to a multivariate analysis of covariance (see Appendix, Table 12). Mother's educational level was designated as the covariate. After controlling for the impact of mother's education, there was still a significant difference

between the ethnic groups for Verbal IQ [$F = 16.347$, $p < .01$] and Full Scale IQ [$F = 18.323$, $p < .01$].

Furthermore, an analysis of covariance was used to evaluate the effects of ethnicity when mother's education and income are held constant as the covariates. Controlling for both of these factors does not eliminate the differences among the three ethnic groups. As shown in Table 13, there was still a significant difference for ethnicity even when mother's education and income were both used as covariates.

Table 14 displays the analysis of variance comparing the impact of the primary diagnosis of the groups on the overall IQ measures. In all groups, the largest percentage of individuals were diagnosed with Attention Deficit/Hyperactivity Disorder (ADHD) (see Appendix, Table 15). African American, Caucasian, and Hispanic males showed no significant differences on the three IQ scales based on primary diagnoses. African American females showed significant differences only on Performance IQ, Caucasian females showed significant differences on Performance and Full IQ, and Hispanic females showed no significant differences on the IQ scales based on primary diagnosis. An analysis of covariance using the primary diagnosis as the covariate still revealed significant differences between the three ethnic groups on Verbal [$F = 17.986$, $p < .01$] and Full Scale IQ [$F = 20.755$, $p < .01$]. Controlling for the influence of primary diagnosis on Performance IQ revealed no significant differences between the ethnic groups on this IQ scale (see Appendix, Table 16).

In order to control for the significant differences between the groups on the variable of income, the data was subjected to another analysis of covariance in which

income was designated as the covariate. A significant difference continued to exist between the ethnic groups on Verbal IQ [$F = 9.444$, $p < .01$], Performance IQ [$F = 9.283$, $p < .01$], and Full Scale IQ [$F = 11.348$, $p < .01$] even after controlling for the effects of income (see Appendix, Table17).

Finally, as used in research by Taylor and Richards (1991), a multivariate analysis of covariance evaluated the patterns of intellectual performance among the groups. Using Full Scale IQ as the covariate, Tables 18 and 20 revealed that significant differences no longer existed between the three ethnic groups on the Verbal and Performance IQ scores. Adjusted means are given in Tables 19 and 21 for both Verbal IQ and Performance IQ for all three ethnic groups when Full Scale IQ was held constant as the covariate.

In addition, no significant differences existed between the three groups on the eleven subtests when FSIQ is held constant. In fact, an investigation of the adjusted means after controlling for Full Scale IQ indicated substantial differences in the patterns of intellectual performance among the three groups (see Appendix, Table 22) as compared to the actual mean scores for the ethnic groups on the eleven subtests.

CHAPTER 4

DISCUSSION

For many years the debate has continued on the costs and benefits of using standardized intelligence tests with children of various ethnic and cultural groups. Research has come to differing conclusions about whether these tests are able to accurately discern the true intellectual capabilities of children from minority groups. The current study evaluated the WISC-III, comparing the performance of African American, Caucasian, and Hispanic children on various demographic variables, on eleven subtests, and on the three IQ scores.

The goal of this study was to investigate whether differences exist among the three ethnic groups and to determine what factors might contribute to any differences. According to the developers of the WISC-III, this most recent version was extensively revised and edited to create a test that was more balanced and appropriate for children of all ethnic groups. Based on this assumption, it was hypothesized that no differences should be apparent among the three ethnic groups for Verbal IQ, Performance IQ, Full Scale IQ, or any of the subtests scores. Furthermore, it was also hypothesized that no differences among the groups would be apparent due to factors such as gender, age, family income, parental educational levels, or diagnosis.

In the present study, test scores of 241 African American, Caucasian, and Hispanic males and females between the ages of seven and sixteen were analyzed. Significant

differences among the groups were found on all three IQ scores and all subtests excluding digit span and coding. Caucasian children obtained significantly higher IQs when compared with the other children. On all subtests except coding, the Caucasian children's mean scores were significantly higher than the two other ethnic groups. On coding, the Hispanic children's mean score was higher, but it was not significantly different from the other two groups. Despite predictions that differences should be minimized on the WISC-III, initial analysis of this clinical sample of children revealed significant differences among the three ethnic groups.

Comparisons of the three ethnic groups divided by gender also revealed significant differences on many of the scores. All three of the ethnic groups for females demonstrated significant differences in Verbal, Performance, and Full Scale IQ. Only Verbal and Full Scale IQ scores were significant among the three ethnic groups for males. Likewise, differences were found among the males for all three groups on every subtest excluding arithmetic, coding, and digit span. The three groups of females were significantly different on all subtest excluding arithmetic, coding, digit span, and picture arrangement. However, there was no evidence of overall gender differences in IQ when comparing the sample as a whole.

As a result of these differences among the ethnic groups, further analysis was employed to determine exactly what factors might contribute to such differences. Comparisons of the demographic variables of the sample produced a mixture of results. No significant differences were found when comparing the mean ages of the three groups as well as no significant differences were found in the primary diagnoses of the individuals

in this study. Most children were diagnosed with ADHD (DSM-IV codes of either 314.00 or 314.01). A significant difference, however, was evident among the groups for family income. African American children had the lowest mean income with Hispanic children next and then Caucasian children.

Investigation of the impact of the demographic variables on test performance produced many various results. It was found that father's educational level had only minimal influence on the differences among the three groups. Only the scores of Caucasian males indicated any relationship between the level of father's education and overall intellectual abilities. For this group, higher educational achievement by the father seemed to be associated with higher IQ scores, particularly with Full Scale IQ.

The lack of influence of father's education on the other groups, particularly the other male groups, may be in part a factor of the many missing variables within the other two ethnic groups. Both African American and Hispanic males had a substantial portion of each group which did not have a father's educational level listed in their files. Many of these children were from homes where only a single mother was present. It is hard to determine exactly what influence the presence of a father might have on the correlation between father's education and these two groups' IQ scores. These two groups might have shown similar patterns of increasing IQ scores with increasing educational attainment had the fathers been present in the home.

Significant correlations were found between mother's educational level and overall IQ scores with all three groups demonstrating a significant relationship between the level of the mother's education and a child's three IQ scores. Gender differences were also

evident within the groups. When split by gender, only African American and Caucasian males' scores evidenced significant differences on the basis of mother's educational level. African American boys exhibited significant differences on both Verbal and Full Scale IQ, while Caucasian boys exhibited significant differences on all three IQ measures. Mother's educational level seemed to positively influence the intellectual performance of these two groups with higher levels of education more associated with higher IQ scores.

In order to determine the magnitude of influence of the mother's education on the initial IQ differences found among the three groups, the influence of this factor was controlled by covariate analysis. Differences were still found to exist among the three ethnic groups. Regardless of mother's educational attainment, differences still existed among the three groups on Verbal IQ and Full Scale IQ. Similarly, differences among the ethnic groups were still apparent when both income and mother's education were held constant as the covariates. Another factor or factors was contributing to the overall differences among these three groups.

As stated, the greatest percentage of each group was categorized as individuals with ADHD as the primary diagnosis. A few other individuals were diagnosed as either Oppositional Defiant Disorder or some sort of learning disability. The results seemed to suggest that the groups were as homogenous as possible in regards to primary diagnosis. When the groups were compared to determine the relationship between diagnosis and overall IQ, there was no significant differences in Verbal, Performance, or Full Scale IQ scores among males on the basis of primary diagnosis. African American females, however, showed significant differences on Performance IQ, and Caucasian females

showed significant differences on Performance and Full Scale IQ scores. Hispanic female scores indicated no significant differences on any of the three IQ scales. Controlling for the influence of diagnosis with a covariate analysis indicated a continued difference among the ethnic groups on Verbal and Full Scale IQ. Diagnosis did not seem to be the discriminating factor among the three groups.

Finally, a significant difference continued to exist among the ethnic groups even after controlling for the effects of income. There was a significant difference among the mean income for the three groups. Although these income differences were held constant, significant differences among the three groups were still evident. Income did not appear to be the underlying factor influencing the apparent differences among the three groups. Theoretically, income should make a large difference in the ability of a family to provide educational materials to enrich the development of a child's intellectual abilities. Income should also influence such basic factors as the availability of libraries, computers, and other resources, the ability of a family to place children in private schools, or the availability of parents to read with their children and participate in their education. However, the effect of income on the scores of these three ethnic groups was not the determining factor in the differences among these three groups.

Using a model taken from Taylor et al., (1991), the differences among scores were evaluated by controlling in part for the difference in Full Scale IQ. Initial analyses indicated that the Caucasian children had the highest actual subtest scores; however, after adjusting the subtest scores using FSIQ as the covariate, different patterns of intellectual abilities emerged. There were no longer any significant differences among the ethnic

groups on Verbal IQ, Performance IQ or the eleven subtest scores. In fact, the adjusted means revealed that the performance of the other two ethnic groups was now higher than the group of Caucasian children on many of the subtests. Of those subtests, Hispanic children performed highest of the three groups on Vocabulary, Digit Span, and Block Design. African American children performed highest of the three groups on Arithmetic, Comprehension, Coding, and Picture Arrangement, while Caucasian children's scores were highest on the remaining four subtests. These results seem to support arguments by Spearman (1927) that the differences among ethnic groups are more related to g, the general intelligence factor, rather than to narrower, more specific factors. When other variables were factored out, only controlling for the Full Scale IQ could eliminate the differences in Verbal IQ scores, Performance IQ scores, and subtest scores of the three ethnic groups in this sample.

Controlling for g also seemed to clearly delineate a pattern of intellectual abilities in which all of the ethnic groups showed varying strengths and weaknesses. African American children seemed to possess relative strengths in tasks requiring planning ability and sequencing. They also seemed to perform better on tasks which used common sense about social judgments. Hispanic children, on the other hand, demonstrated good conceptualizing abilities with better performance also on tasks requiring memory and attention. Finally, Caucasian children seemed to perform better on abstract tasks which used verbal concept formation and associative thinking. They also exhibited good verbal comprehension and understanding of relationships between objects.

All of these results suggest a number of important implications for the future of intelligence and educational testing for children. Most importantly, these results suggest that the Wechsler Intelligence Scale for Children-Third Edition is not biased towards minority groups. When variations in Full Scale IQ are taken into account, minority children perform as well or better than Caucasian children on many of the subtests. Investigation of the patterns of intellectual abilities which emerge when Full Scale IQ is held constant suggest that the WISC-III does accurately reflect the abilities of minority children.

These findings underline the necessity of considering many different factors when testing children for placement in learning disabled and special education programs. Accepting a single standardized IQ score at face value can lead to erroneous conclusions. At first glance, differences in IQ scores among various ethnic groups seem quite large. As a result, children of different ethnic groups may be more likely to be placed in special education programs because their scores will appear significantly lower than their Caucasian peers. However, when the influence of g is taken into account, a more accurate picture of a child's abilities can be discovered.

As found in recent literature on the WISC-III, IQ scores do tend to decrease as a product of norming and may be more likely to disqualify children as learning disabled and to place children in special education programs. Since this current study showed that minority children's scores may look significantly lower than Caucasian children, they may be more likely to be placed in these special education classes in disproportionate amounts as compared to their Caucasian peers. As with any revised version of an older test, this

shift in scores is predictable. However, the apparent decrease may be more disproportionate for minorities, suggesting that thorough evaluations of a child's pattern of abilities is important in assessing the educational program that is best suited for that child. The WISC-III seems to be a valid and useful tool for evaluating the abilities of children if less emphasis is placed on the specific IQ scores and more emphasis is placed on the pattern of intellectual abilities of each individual child.

The results of this study place more work on the clinician/examiner to administer a host of tests which evaluate many aspects of a child's abilities, to carefully examine test results for patterns of abilities, and to not presume that children with the same overall IQ necessarily have the same pattern of strengths and weaknesses. These results suggest the need to decrease the importance on single overall IQ scores and specific IQ score cut-offs for placement and diagnosis of children with disabilities. In agreement with Steve Edelman (1996), the results of this study suggest that school systems which still rely on ability test-achievement test score discrepancies may be detrimentally using such test results for placement and re-evaluation of exceptional children.

There a few limitations to this study which should be considered and further evaluated. First, there was no way to determine the level of proficiency in the English language for the Hispanic children. All Hispanic children were administered the WISC-III in English, and how this administration of a test in a language other than an individual's primary language has effected the overall results of the study is unknown. However, the Hispanic children, interestingly, received the highest adjusted mean subtest score on the

subtest, Vocabulary. This subtest has been defined as measuring word knowledge, verbal comprehension, and language development (Sattler, 1992).

Secondly, it is possible that the differences among the groups were also influenced by the diagnoses other than the primary diagnosis. Most of the children had more than one diagnosis. This study looked only at the primary diagnosis with the assumption that this diagnosis was most likely the focus of the child's problems and treatment. In addition, this study did not investigate which children were prescribed psychotropic medications for their respective diagnosis. No comments can be made on how such medications might impact performance on an intelligence test.

Thirdly, in the analyses of parental education, several of the cells had very small sample sizes (N). The small number of cases in some of the cells could potentially decrease the power of the significance in these evaluations. However, this limitation does not impact the overall findings of this study which found that no differences existed among the three ethnic groups only when Full Scale IQ was held constant. The factors of parent education were significant for some groups; however, any significance based on parent education was held constant using an analysis of covariance. These analyses demonstrated that differences continued to exist among the three ethnic groups regardless of the father's or mother's educational level.

Finally, the information in this study was obtained through retrieval of archival data from the files of the subjects included in the study. The results could be influenced by an examiner effect not addressed in the study. However, all of the individuals were administered the WISC-III by one of three individuals, and the individual test protocols

were examined by the primary author of this study in an attempt to avoid arithmetic and scoring errors.

Future research should focus on discovering alternatives to global IQ scores. Finding new ways to characterize a child's abilities seems to be an important advancement in the field of intelligence testing. Can a child's abilities be accurately reflected without assigning a specific IQ score? Furthermore, it seems likely that there still exists some environmental factors which influence the apparent differences seen in the scores of minority and Caucasian children. The influence of cultural values and language differences on intellectual development seems to be an area of research which is not only difficult to observe but also difficult to document and analyze. Until a new approach to intelligence testing for children of various ethnicities can be developed, the WISC-III provides one of the best evaluations of a child's abilities. With thorough investigation of a child's pattern of intellectual abilities within the WISC-III, an examiner can accurately assess the educational and intellectual needs of children from all cultural and ethnic groups.

APPENDIX

TABLES

Table 1

Review of the Literature on the WISC-R and WISC-III

WISC-R Research	WISC-III Research
1. Factor Structure Geary & Whitworth (1988) Carlson, Reynolds, & Gutkin (1983) Dean (1980) Johnston & Bolen (1984) Oakland & Feigenbaum (1979)	1. Factor Structure-- A. Support Four Factor Wechsler (1991) Roid, Prifitera, & Weiss (1993) Hishinuma & Yamakawa (1993) B. Use three factor, Four Factor needs more research Sattler (1992) Kamphaus et al., (1994) Reynolds & Ford (1994) Slate & Jones (1995)
2. Test Bias Among Minority Groups A. No substantial bias Reynolds & Rechsley (1983) Sandoval (1979) Vance, Hankins, & McGee (1979) PASE case (1980) Marshall v. Georgia (1984) Taylor & Richards (1996) Rechsley & Rechsley (1979) Oakland (1980) Dean (1979) B. Significant bias Munford (1978) Arinoldo (1981) Munford & Munoz (1980) Reynolds & Gutkin (1981) Munford, Meyerowitz, & Munford (1980) Larry P. case (1972) Diana case (1970) Hobson v Hansen (1967) C. Alternatives to standardized tests Willliams (1972) Cattell (1959) Raven (1965) Leiter (1969) Shimberg (1929) Quay (1971)	2. Criterion Validity Sattler & Atkinson (1993)-pos. correl. with WPPSI-R Prewett & Matavich (1994) correl. with SB:IV Lukens & Hurrell (1996) correl. with SB:IV Lavin (1996) correl. with SB:IV Rust & Yates (1997) correl. with SB:IV Slate et al., (1994) correl. with PPVT-R Slate (1995) correl. with PPVT-R Rust & Yates (1997) correl with K-ABC 3. WISC-R to WISC-III Smith et al., (1994) decreases/hi. correl. Dumont & Faro (1993) decreases Hagar (1992) greater decrease for Afr. Am. Nagle & Daley (1994) need more research Lyon (1995) predicted decreases Slate (1995) decreases Bolen et al., (1995) decreases Gunter et al., (1995) decreases Slate & Saarnio (1995) decreases Graf & Hinton (1994) decreases but hi correl. Slate et al., (1995) some increase in scores Ewing (1994) increase

Table 2

Demographic Variables

	N	Mean Age	Mean Income
African American	70	11.31	$10,733 **
Males	35	10.52	$ 9,594
Females	35	12.08	$11,872 *
Caucasian	102	11.92	$24,971**
Males	51	11.56	$27,631
Females	51	12.28	$22,311
Hispanic	69	10.80	$17,723**
Males	35	9.98	$13,959
Females	34	11.62	$21,597**

* $p < .05$
** $p < .01$

Table 3

Anova of Gender Differences on WISC-III IQ and Subtest Scores

Male (N=121)	Mean	Female (N=120)	Mean
VIQ	90.74	VIQ	90.87
PIQ	91.69	PIQ	91.22
FSIQ	90.41	FSIQ	90.23
INFO	8.07	INFO	7.83
SIM	8.83	SIM	8.57
ARITH	7.86	ARITH	8.09
VOCAB	8.01	VOCAB	8.23
COMPR	8.84	COMPR	8.97
DIGIT SP	8.05	DIGIT SP	8.04
PIC COMP	8.97	PIC COMP	8.54
COD	7.93	COD	9.09**
PIC ARR	8.26	PIC ARR	8.37
BLK DES	8.61	BLK DES	8.33
OBJ ASSEM	9.21*	OBJ ASSEM	8.34

* $p < .05$
** $p < .01$

Table 4

Anova of Differences Among Ethnic Groups on Overall IQ Measures

		N	Mean	df	F	Sig.
VIQ	Afr. Am.	70	83.871	2	17.369	.000
	Cauc.	102	96.088			
	Hisp.	69	90.029			
PIQ	Afr. Am.	70	83.871	2	16.444	.000
	Cauc.	102	96.833			
	Hisp.	69	91.217			
FSIQ	Afr. Am.	70	82.471	2	20.111	.000
	Cauc.	102	96.118			
	Hisp.	69	89.725			

Table 5

Anova of Differences Among Ethnic Groups on WISC-III Subtests

		N	Mean	df	F	Sig.
INFO	Afr. Am.	70	6.786	2	13.848	.000
	Cauc.	102	9.039			
	Hisp.	69	7.536			
SIM	Afr. Am.	70	7.214	2	15.807	.000
	Cauc.	102	9.833			
	Hisp.	69	8.551			
ARITH	Afr. Am.	70	7.471	2	3.093	.047
	Cauc.	102	8.402			
	Hisp.	69	7.870			
VOCAB	Afr. Am.	70	6.500	2	18.338	.000
	Cauc.	102	9.088			
	Hisp.	69	8.333			
COMPR	Afr. Am.	70	7.629	2	10.413	.000
	Cauc.	102	9.853			
	Hisp.	69	8.812			
DIGIT SP	Afr. Am.	70	7.400	2	3.018	.051
	Cauc.	102	8.473			
	Hisp.	69	8.092			
PIC COMP	Afr. Am.	70	7.200	2	19.776	.000
	Cauc.	102	10.000			
	Hisp.	69	8.493			
COD	Afr. Am.	70	8.200	2	.536	.586
	Cauc.	102	8.628			
	Hisp.	69	8.652			
PIC ARR	Afr. Am.	70	7.243	2	6.838	.001
	Cauc.	102	9.196			
	Hisp.	69	8.101			
BLK DES	Afr. Am.	70	6.814	2	13.023	.000
	Cauc.	102	9.441			
	Hisp.	69	8.725			
OBJ ASSEM	Afr. Am.	70	7.371	2	11.469	.000
	Cauc.	102	9.775			
	Hisp.	69	8.725			

Table 6

Differences on Overall IQ and Subtest Scores Among Ethnic Groups Split By Gender

Male	df	F	Sig.	Female	df	F	Sig.
VIQ	2	9.34	.000	VIQ	2	8.04	.001
PIQ	2	1.9	.153	PIQ	2	5.92	.004
FSIQ	2	12.51	.000	FSIQ	2	7.83	.001
INFO	2	6.39	.002	INFO	2	7.32	.001
SIM	2	11.3	.000	SIM	2	5.33	.006
ARITH	2	1.85	.160	ARITH	2	2.85	.062
VOCAB	2	10.62	.000	VOCAB	2	7.99	.001
COMPR	2	5.15	.007	COMPR	2	5.18	.007
DIGIT SP	2	0.71	.490	DIGIT SP	2	2.67	.073
PIC COMP	2	17.66	.000	PIC COMP	2	4.58	.012
COD	2	1.02	.362	COD	2	2.3	.105
PIC ARR	2	7.56	.001	PIC ARR	2	1.8	.169
BLK DES	2	6.43	.002	BLK DES	2	6.5	.002
OBJ ASSEM	2	6.74	.002	OBJ ASSEM	2	5.1	0.01

Males: Afr. Am. N = 35, Cauc. N = 51, Hisp. N = 35
Females: Afr. Am. N = 35, Cauc. N = 51, Hisp. N = 34

Table 7

Anova of Differences in IQ Scores for Males Based on Dad's Education Level

		N	Mean	df	F	Sig.
Afr. Am. VIQ	<= 8th	1	62.000	4	1.958	0.147
	9th - 11th	3	90.667			
	12th	14	88.571			
	13-15	3	86.667			
	16+	1	106.000			
Afr. Am. PIQ	<= 8th	1	90.000	4	0.695	0.605
	9th - 11th	3	91.000			
	12th	14	85.928			
	13-15	3	86.667			
	16+	1	104.000			
Afr. Am. FSIQ	<= 8th	1	73.000	4	1.373	0.285
	9th - 11th	3	89.667			
	12th	14	86.071			
	13-15	3	85.667			
	16+	1	106.000			
Cauc. VIQ	<= 8th	1	69.000	4	1.717	0.165
	9th - 11th	7	93.000			
	12th	12	94.9167			
	13-15	19	97.6842			
	16+	6	103.167			
Cauc. PIQ	<= 8th	1	54.000	4	3.407	0.017
	9th - 11th	7	95.1429			
	12th	12	102.167			
	13-15	19	99.368			
	16+	6	106.833			
Cauc. FSIQ	<= 8th	1	58.000	4	2.881	0.035
	9th - 11th	7	93.142			
	12th	12	98.000			
	13-15	19	98.736			
	16+	6	105.000			
Hisp. VIQ	<= 8th	6	88.167	4	0.689	0.607
	9th - 11th	11	86.545			
	12th	9	95.111			
	13-15	2	92.000			
	16+	1	100.000			
Hisp. PIQ	<= 8th	6	84.167	4	0.709	0.594
	9th - 11th	11	88.454			
	12th	9	88.222			
	13-15	2	107.500			
	16+	1	83.000			
Hisp. FSIQ	<= 8th	6	85.000	4	0.479	0.751
	9th - 11th	11	86.181			
	12th	9	91.111			
	13-15	2	99.500			
	16+	1	92.000			

Table 8

Anova of Differences In IQ Scores For Females Based on Dad's Education Level

		N	Mean	df	F	Sig.
Afr. Am. VIQ	<= 8th	1	72.000	4	0.285	0.884
	9th - 11th	5	88.200			
	12th	13	86.461			
	13-15	3	86.333			
	16+	1	87.000			
Afr. Am. PIQ	<= 8th	1	86.000	4	0.085	0.986
	9th - 11th	5	83.800			
	12th	13	88.077			
	13-15	3	84.667			
	16+	1	83.000			
Afr. Am. FSIQ	<= 8th	1	77.000	4	0.085	0.986
	9th - 11th	5	84.600			
	12th	13	86.076			
	13-15	3	84.000			
	16+	1	84.000			
Cauc. VIQ	<= 8th	3	94.333	4	0.264	0.899
	9th - 11th	6	93.500			
	12th	17	95.823			
	13-15	12	96.833			
	16+	5	102.000			
Cauc. PIQ	<= 8th	3	93.333	4	1.139	0.353
	9th - 11th	6	91.000			
	12th	17	91.588			
	13-15	12	101.083			
	16+	5	101.200			
Cauc. FSIQ	<= 8th	3	93.333	4	0.607	0.66
	9th - 11th	6	91.667			
	12th	17	93.176			
	13-15	12	98.917			
	16+	5	101.800			
Hisp. VIQ	<= 8th	6	84.333	4	2.281	0.09
	9th - 11th	8	83.500			
	12th	8	101.625			
	13-15	4	91.000			
	16+	3	90.333			
Hisp. PIQ	<= 8th	6	84.000	4	1.806	0.161
	9th - 11th	8	92.875			
	12th	8	103.250			
	13-15	4	96.500			
	16+	3	86.333			
Hisp. FSIQ	<= 8th	6	83.167	4	2.105	0.111
	9th - 11th	8	86.750			
	12th	8	102.375			
	13-15	4	92.750			
	16+	3	87.333			

Table 9

Differences in Overall IQ Scores Among Ethnic Groups Based on Mom's Education Level

		N		df	F	Sig.
Afr Am	<= 8th	3	VIQ	4	4.405	0.003
	9th - 11th	15				
	12th	36				
	13-15	14				
	16+	1				
	<= 8th	3	PIQ	4	2.027	0.101
	9th - 11th	15				
	12th	36				
	13-15	14				
	16+	1				
	<= 8th	3	FSIQ	4	3.096	0.022
	9th - 11th	15				
	12th	36				
	13-15	14				
	16+	1				
Cauc	<= 8th	5	VIQ	4	6.088	0.000
	9th - 11th	17				
	12th	43				
	13-15	24				
	16+	11				
	<= 8th	5	PIQ	4	5.484	0.001
	9th - 11th	17				
	12th	43				
	13-15	24				
	16+	11				
	<= 8th	5	FSIQ	4	7.33	0.000
	9th - 11th	17				
	12th	43				
	13-15	24				
	16+	11				
Hisp	<= 8th	12	VIQ	4	3.522	0.012
	9th - 11th	14				
	12th	25				
	13-15	14				
	16+	4				
	<= 8th	12	PIQ	4	3.08	0.022
	9th - 11th	14				
	12th	25				
	13-15	14				
	16+	4				
	<= 8th	12	FSIQ	4	3.772	0.008
	9th - 11th	14				
	12th	25				
	13-15	14				
	16+	4				

Table 10

Mom's Education Level By Ethnicity And Gender

		Male (N)	Female (N)
African American	<= 8th	1	2
	9th - 11th	6	9
	12th	18	18
	13-15	9	5
	16+	0	1
Caucasian	<= 8th	2	3
	9th - 11th	9	8
	12th	18	25
	13-15	14	10
	16+	7	4
Hispanic	<= 8th	6	6
	9th - 11th	10	4
	12th	11	14
	13-15	6	8
	16+	2	2

Table 11

Relationship of Mom's Education To Overall IQ Scores Among Ethnic Groups Separated by Gender

	Male	df	F	Sig.	Female	df	F	Sig.
Afr Am	VIQ	3	4.740	.008	VIQ	4	1.575	.206
	PIQ	3	.952	.428	PIQ	4	1.468	.237
	FSIQ	3	2.916	.050	FSIQ	4	1.223	.322
Cauc	VIQ	4	5.138	.002	VIQ	4	1.778	.150
	PIQ	4	4.965	.002	PIQ	4	2.465	.058
	FSIQ	4	6.190	.000	FSIQ	4	2.384	.065
Hisp	VIQ	4	2.406	.071	VIQ	4	1.943	.130
	PIQ	4	1.344	.277	PIQ	4	2.338	.079
	FSIQ	4	2.11	.104	FSIQ	4	2.367	.076

Table 12

Analysis of Covariance With Mother's Education as the Covariate

	Source	df	F	Sig.
Dependent Var.--VIQ	Mom's Ed.	1	35.951	0.000
	Ethnicity	2	16.347	0.000
Dependent Var.--PIQ	Mom's Ed	1	0.004	0.948
	Ethnicity	2	2.841	0.06
Dependent Var.--FSIQ	Mom's Ed	1	35.722	0.000
	Ethnicity	2	18.323	0.000

Table 13

Analysis of Covariance With Mom's Education and Income as the Covariates

	Source	df	F	Sig.
Dependent Var.--VIQ	Mom's Ed.	1	23.483	0.000
	Income	1	2.104	0.148
	Ethnicity	2	11.446	0.000
Dependent Var.--PIQ	Mom's Ed.	1	0.025	0.874
	Income	1	0.051	0.822
	Ethnicity	2	2.377	0.095
Dependent Var.--FSIQ	Mom's Ed	1	21.821	0.000
	Income	1	2.615	0.107
	Ethnicity	2	13.327	0.000

Table 14

Effect of Primary Diagnosis on Overall IQ Scores

	Male	df	F	Sig.	Female	df	F	Sig.
Afr Am	VIQ	5	.180	.968	VIQ	10	1.567	.177
	PIQ	5	.527	.754	PIQ	10	2.744	.021
	FSIQ	5	.296	.911	FSIQ	10	2.123	.064
Cauc	VIQ	9	1.446	.202	VIQ	11	1.525	.162
	PIQ	9	1.192	.327	PIQ	11	2.409	.021
	FSIQ	9	1.430	.208	FSIQ	11	2.071	.047
Hisp	VIQ	11	.756	.677	VIQ	10	1.411	.237
	PIQ	11	.921	.538	PIQ	10	.977	.489
	FSIQ	11	.846	.600	FSIQ	10	1.180	.352

Table 15

Number of Diagnoses

	Male		Female	
	Diagnosis	Number	Diagnosis	Number
African American	61.80	1	61.20	1
	62.81	1	62.89	1
	313.81	1	313.81	1
	314.00	4	314.00	5
	314.01	27	314.01	24
	315.00	1	315.20	3
Caucasian	61.20	2	61.20	2
	61.80	2	61.80	2
	62.30	1	62.30	1
	307.23	1	300.30	1
	313.81	1	307.23	1
	314.00	5	309.24	2
	314.01	37	313.81	2
	315.00	2	314.00	6
			314.01	34
Hispanic	61.80	1	61.20	1
	62.30	1	62.30	2
	62.81	1	62.89	1
	307.23	2	314.00	6
	313.81	3	314.01	22
	314.00	2	315.10	1
	314.01	23	315.31	1
	314.10	1		
	315.20	1		

Table 16

Analysis of Covariance with Primary Diagnosis as the Covariate

	Source	df	F	Sig.
Dependent Var.--VIQ	Diagnosis	1	4.463	0.036
	Ethnicity	2	17.986	0.000
Dependent Var.--PIQ	Diagnosis	1	0.335	0.563
	Ethnicity	2	2.951	0.054
Dependent Var.--FSIQ	Diagnosis	1	4.041	0.046
	Ethnicity	2	20.755	0.000

Table 17

Analysis of Covariance with Income as the Covariate

	Source	df	F	Sig.
Dependent Var.--VIQ	Income	1	15.000	0.000
	Ethnicity	2	9.444	0.000
Dependent Var.--PIQ	Income	1	12.007	0.001
	Ethnicity	2	9.283	0.000
Dependent Var.--FSIQ	Income	1	15.526	0.000
	Ethnicity	2	11.348	0.000

Table 18

Analysis of Covariance with FSIQ as the Covariate

Dependent Variable: VIQ

Source	df	F	Sig.
FSIQ	1	12.846	.000
ETHNICITY	2	0.115	.891

Table 19

Adjusted Means for Verbal IQ

Dependent Variable: VIQ			95% Confidence Interval	
Ethnicity	Mean	Std. Error	Lower Bound	Upper Bound
Afr Am	90.807*	.680	89.468	92.146
Cauc	90.971*	.558	89.872	92.070
Hisp	90.558*	.655	89.268	91.848

*Evaluated at covariates appeared in the model: FSIQ = 90.237

Table 20

Analysis of Covariance for Performance IQ

Dependent Variable: PIQ

Source	df	F	Sig.
FSIQ	1	6.783	.010
ETHNICITY	2	0.879	.416

Table 21

Adjusted Means for Performance IQ

Dependent Variable: PIQ			95% Confidence Interval	
Ethnicity	Mean	Std. Error	Lower Bound	Upper Bound
Afr Am	89.544*	7.433	74.901	104.188
Cauc	101.471*	6.103	89.448	113.494
Hisp	91.650*	7.160	77.545	105.755

* Evaluated at covariates appeared in the model: FSIQ = 90.3237

Table 22

Subtest Means and Adjusted Means

		Mean	Adjusted Mean
INFO	AFR. AM.	6.7857	7.9910
	CAUC.	9.0392	8.1500
	HISP.	7.5362	7.6280
SIM	AFR. AM.	7.2143	8.5660
	CAUC.	9.8333	8.8360
	HISP.	8.5507	8.6540
ARITH	AFR. AM.	7.4714	8.3990
	CAUC.	8.4020	7.7180
	HISP.	7.8696	7.9400
VOCAB	AFR. AM.	6.5000	7.7300
	CAUC.	9.0882	8.1810
	HISP.	8.3333	8.4270
COMPR	AFR. AM.	7.6286	9.0070
	CAUC.	9.8529	8.8360
	HISP.	8.8116	8.9170
DIGIT SP	AFR. AM.	7.4000	7.8940
	CAUC.	8.4731	8.0810
	HISP.	8.0923	8.1580
PIC COMP	AFR. AM.	7.2000	8.2650
	CAUC.	10.0000	9.2140
	HISP.	8.4928	8.5740
COD	AFR. AM.	8.2000	9.0500
	CAUC.	8.6275	8.0010
	HISP.	8.6522	8.7170
PIC ARR	AFR. AM.	7.2429	8.7070
	CAUC.	9.1961	8.1160
	HISP.	8.1014	8.2130
BLK DES	AFR. AM.	6.8143	8.1460
	CAUC.	9.4412	8.4580
	HISP.	8.7246	8.8260
OBJ ASSEM	AFR. AM.	7.3714	8.6150
	CAUC.	9.7745	8.8570
	HISP.	8.7246	8.8190

REFERENCES

Angoff, W.H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R.A. Berk (Ed), Handbook of methods for detecting test bias (pp. 96-116). Baltimore, MD: The Johns Hopkins University Press.

Arinoldo, C.G. (1981). Black-white differences in the general cognitive index of the McCarthy scales and in the full scale IQs of Wechsler's scales. Journal of Clinical Psychology, 37, 630-637.

Binet, A. & Simon, T. (1916). The development of intelligence in children Trans. by Elizabeth S. Kite. Baltimore: Williams & Wilkins.

Block, N.J. & Dworkin, G. eds. (1976). The IQ controversy. New York: Pantheon Books.

Blumberg, T.A. (1995). A practitioner's view of the WISC-III. Journal of School Psychology, 33, 95-97.

Bolen, L.M., Aichinger, K.S., Hall, C.W., & Webster, R.E. (1995). A comparison of the performance of cognitively disabled children on the WISC-R and WISC-III. Journal of Clinical Psychology, 51, 89-94.

Burt, C. (1966). The genetic determination of differences in intelligence: A study of monozygotic twins reared together and apart. British Journal of Psychology, 57, 137-153.

Caravajal, H., Hayes, J.E., Miller, H.R., Wiebe, D.A., & Weaver, K.A. (1993). Comparisons of the Vocabulary scores and IQs on the Wechsler Intelligence Scale for Children-III and the Peabody Picture Vocabulary Test-Revised. Perceptual and Motor Skills, 76, 28-30.

Carlson, L.C., Reynolds, C.R., & Gutkin, T.B. (1983). Consistency of the factorial validity of the WISC-R for upper and lower SES groups. Journal of School Psychology, 21, 319-326.

Cattell, R.B. (1971). Abilities: Their structure, growth, and action. Boston: Houghton Mifflin Company.

Cattell, R.B. & Cattell, A.K.S. (1959). Handbook for the culture fair intelligence test. Champaign, Ill.: Institute for Personality and Ability Testing.

Cleary, A.T. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 5, 115-124.

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

Dean, R.S. (1979). Predictive validity of the WISC-R with Mexican American children. Journal of School Psychology, 17, 55-58.

Dean, R.S. (1980). Factor structure of the WISC-R with Anglos and Mexican Americans. Journal of School Psychology, 18, 234-239.

DuBois, P.H. (1970). A history of psychological testing. Boston: Allyn and Bacon, Inc.

Dumont, R., & Faro, C. (1993). The WISC-III: Almost two years old; proceeding with caution--practitioners' concerns. Communique, 21(7), 12-15.

Edelman, S. (1996). A review of the Wechsler Intelligence Scale for Children-Third Edition (WISC-III). Measurement and Evaluation in Counseling and Development, 28, 219-224.

Edwards, A.J. (1971). Individual mental testing. San Francisco: Intext Educational Publishers.

Ewing, M.L. (1994). The effects of cultural bias: A comparison of the WISC-R and the WISC-III. An unpublished Master's thesis, University of North Texas, Denton, Texas.

Eysenck, H.J. (1971). The IQ argument. New York: McGraw-Hill Publishing Company.

Eysenck, H.J. & Kamin, L. (1981). The intelligence controversy. New York: John Wiley and Sons.

Geary, D.C., & Whitworth, R.H. (1988). Is the factor structure of the WISC-R different for Anglo-and Mexican-American children? Journal of Psychoeducational Assessment, 6, 253-260.

Goddard, H.H. (1908). The Binet and Simon tests of intellectual capacity. Training School, 5, 3-9.

Good, W.R. (1954). Misconceptions about intelligence testing. The Education Digest, 20, 14-16.

Graf, M.H., & Hinton, R.N. (1994). A 3-year comparison study of WISC-R and WISC-III IQ scores for a sample of special education students. Educational and Psychological Measurement, 43(1), 128-133.

Gunter, C.M., Sapp, G.L., & Green, A.C. (1995). Comparison of scores on WISC-III and WISC-R of urban learning disabled students. Psychological Reports, 77, 473-474.

Hager, V. (1992). A study of WISC-II scores in 11- to 14-year-old learning disabled children. Unpublished educational specialist thesis, James Madison University, Harrisonburg, VA.

Heathers, G. (1969). Encyclopedia of educational research. 4th ed. Ed. R.L. Ebel. New York: The Macmillan Company, 559-570.

Hishinuma, E.S., & Yamakawa, R. (1993). Construct and criterion-related validity of the WISC-III for exceptional students and those who are "at risk". Journal of Psychoeducational Assessment: WISC-III Monograph, 94-104.

Hothersall, D. (1990). History of psychology. New York: McGraw-Hill Publishing Company.

Husband, T.H., & Hayden, D.C. (1996). Effects of the addition of color to assessment instruments. Journal of Psychoeducational Assessment, 14,147-151.

Jensen, A.R. (1969). How much can we booster IQ and scholastic achievement? Harvard Educational Review, 39.

Jensen, A.R. (1980). Bias in mental testing. New York: The Free Press.

Johnston, W.T., & Bolen, L.M. (1984). A comparison of the factor structure of the WISC-R for blacks and whites. Psychology in Schools, 21, 42-44.

Joseph, A. (1977). Intelligence, IQ and race--When, how and why they became associated. San Francisco: R & E Research Associates, Inc.

Kamphaus R.W. (1993). Clinical assessment of children's intelligence. Boston: Allyn and Bacon.

Kamphaus, R.W, Benson, J., Hutchinson, S., & Platt, L.O. (1994). Identification of factor models for the WISC-III. Educational and Psychological Measurement, 54(1), 174-186.

Kamphaus, R.W. & Platt, L.O. (1992). Subtests specificities for the WISC-III. Psychological Reports, 70, 899-902.

Kaufman, A.S. (1979). Intelligent testing with the WISC-R. New York: Wiley-Interscience.

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

Kaufman, A.S. (1993). King WISC the Third assumes the throne. The Journal of School Psychology, 31, 345-354.

Kaufman, A.S.. McLean, J.E., & Reynolds, C.R. (1988). Sex, race, residence, region, and education differences on the 11 WAIS-R subtests. Journal of Clinical Psychology, 44, 231-248.

Kramer, J.H. (1993). Interpretation of individual subtest scores on the WISC-III. Psychological Assessment, 5, 193-196.

Lavin, C. (1996). The Wechsler Intelligence Scale for Children-Third Edition and the Stanford-Binet Intelligence Scale: Fourth Edition: a preliminary study of validity. Psychological Reports, 78, 491-496.

Leiter, R.G. (1969). Examiner manual for the Leiter International Performance Scale. Chicago: Stoelting.

Loehlin, J.C., Lindzey, G., & Spuhler, J.N. (1975). Race differences in intelligence. San Francisco: W.H. Freeman and Company.

Lukens, J., & Hurrell, R.M. (1996). A comparison of the Stanford-Binet IV and the WISC-III with mildly retarded children. Psychology in the Schools, 33, 24-27.

Lyon, M.A. (1995). A Comparison between WISC-III and WISC-R scores for learning disabled reevaluations. Journal of Learning Disabilities, 4, 253-255.

Munford, P.R. (1978). A comparison of the WISC and WISC-R on black psychiatric outpatients. Journal of Clinical Psychology, 34, 938-943.

Munford, P.R., Meyerowitz, B.E., & Munford, A.M. (1980). A comparison of black and white children's WISC/WISC-R differences. Journal of Clinical Psychology, 36, 471-475.

Munford, P.R. & Munoz, A. (1980). A comparison of the WISC and WISC-R on Hispanic children. Journal of Clinical Psychology, 36, 452-458.

Nagle, R.J. & Daley, C.E. (1994, March). Longitudinal comparability of the WISC-III and WISC-R with EMR students. Paper presented at the Annual Convention of the National Association of School Psychologists, Seattle, Washington.

Oakland, T. (1980). An evaluation of the ABIC, pluralistic norms, and estimated learning potential. Journal of School Psychology, 18, 3-11.

Oakland, T. & Feigenbaum, D. (1979). Multiple sources of test bias on the WISC-R and Bender-Gestalt Test. Journal of Consulting and Clinical Psychology, 47, 968-974.

Post, K.R. & Mitchell, H.R. (1993). The WISC-III: a reality check. Journal of School Psychology, 31, 541-545.

Prewett, P.N., & Matavich, M.A. (1994). A comparison of referred students' performance on the WISC-III and the Stanford-Binet Intelligence Scale: Fourth Edition. Journal of Psychoeducational Assessment, 12, 42-48.

Quay, L.C. (1971). Language dialect, reinforcement, and the intelligence-test performance of Negro children. Child Development, 42, 5-15.

Raven, J.C. (1965). Guide to the Advanced Progressive Matrices. London: H.K. Lewis.

Rechsly, D., & Rechsly, J. (1979). Validity of the WISC-R factor scores in predicting achievement and attention for four sociocultural groups. Journal of School Psychology, 17, 355-359.

Reynolds, C.R., & Ford, L. (1994). Comparative three-factor solutions of the WISC-III and WISC-R at 11 age levels between 6-1/2 and 16-1/2 years. Archives of Clinical Neuropsychology, 9(6), 553-570.

Reynolds, C.R. & Gutkin, T.B. (1981). A multivariate comparison of the intellectual performance of black and white children matched on four demographic variables. Personality and Individual Differences, 2, 175-180.

Reynolds, C.R. & Reschly, D.J. (1983). An investigation of item bias on the WISC-R with four sociocultural groups. Journal of Consulting and Clinical Psychology, 51, 144-146.

Reynolds, C.R., Willson, V.L. & Jensen, A.R. (1984). Black-white differences in simultaneous and sequential processing independent of overall ability difference. Paper presented at the annual meetings of the American Educational Research Association, New Orleans.

Robb, G.P., Bernardoni, L.C., & Johnson, R.W. (1972). Assessment of individual mental ability. New York: Intext Educational Publishers.

Roid, G.H., Prifitera, A., & Weiss, L.G. (1993). Replication of the WISC-III factor structure in an independent sample. Journal of Psychoeducational Assessment: WISC-III Monograph, 6-21.

Rust, J.O., & Lindstrom, A. (1996). Concurrent validity of the WISC-III and Stanford-Binet IV. Psychological Reports, 79, 618-620.

Rust, J.O., & Yates, A.G. (1997). Concurrent validity of the Wechsler Intelligence Scale for Children-Third Edition and the Kaufman Assessment Battery for Children. Psychological Reports, 80, 89-90.

Samuda, R.J. (1975). Psychological testing of american minorities: Issues and consequences. New York: Dodd, Mead & Company.

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.M. (1992). Assessment of children. San Diego: Jerome M. Sattler, Publisher.

Sattler, J.M. & Atkinson. (1993). Item equivalence across scales: The WPPSI-R and WISC-III. Psychological Assessment, 5, 203-206.

Scarr, S. (1981). Race, social class, and individual differences in IQ. New Jersey: Lawrence Erlbaum Associates.

Shimberg, M.E. (1929). An investigation into the validity of norms with special reference to urban and rural groups. Archives of Psychology, 104.

Slate, J.R. (1995). Discrepancies between IQ and index scores for a clinical sample of students: useful diagnostic indicators? Psychology in the Schools, 32, 103-108.

Slate, J.R. (1995). Two investigations of the validity of the WISC-III. Psychological Reports, 76, 299-306.

Slate, J.R., & Jones, C.H. (1995). Preliminary evidence of the validity of the WISC-III for African American students undergoing special education evaluation. Educational and Psychological Measurement, 55, 1039-1046.

Slate, J.R., Jones, C.H., Graham, L.S., & Gower, J. (1994). Correlations of WISC-III, WRAT-R, KM-R, and PPVT-R scores in students with specific learning disabilities. Learning Disabilities Research and Practice, 9, 104-107.

Slate, J.R., Jones, C.H., & Saarnio, D.A. (1997). WISC-III IQ scores and special education diagnosis. The Journal of Psychology, 131(1), 119-120.

Slate, J.R., & Saarnio, D.A. (1995). Differences between WISC-III and WISc-R IQs: A preliminary investigation. Journal of Psychoeducational Assessment, 13, 340-346.

Smith, D.K., Stovall, D.L., & Geraghty, B. (1994, March). WISC-III/WISC-R relationships in special education re-evaluations. Paper presented at the Annual Meeting of the National Association of School Psychologists, Seattle, Washington.

Spearman, C.E. (1904). General Intelligence: Objectively determined and measured. American Journal of Psychology, 15, 201-292.

Taylor, R.L. & Richards, S.B. (1991). Patterns of intellectual differences of Black, Hispanic, and White children. Psychology in the Schools, 28, 5-9.

Terman, L.M. (1916). The measurement of intelligence. Boston: Houghton-Mifflin.

Truch, S. (1989). The WISC-R companion. Seattle: Special Child Publications.

Vance, H.B., Hankins, N. & McGee, H. (1979). A preliminary study of black and white differences on the revised Wechsler Intelligence Scale for Children. Journal of Clinical Psychology, 35, 815-819.

Vernon, P.E. (1979). Intelligence: Heredity and environment. San Francisco: W.H. Freeman and Company.

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

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

Wechsler, D. (1991). Manual for the Wechsler Intelligence Scale for Children--Third Edition. San Antonio: The Psychological Corporation.

Williams, R.L. (1972) The BITCH-100: A culture-specific test. Paper presented at the meeting of the American Psychological Association, Honolulu, Hawaii.

Wright, B.D. & Stone, M.H. (1979). Best test design. Chicago: MESA Press.

COURT CASES

Brown v. Board of Education of Topeka, Kansas, United States Supreme Court, 347 U.S. 483 (1954).

Diana v. State Board of Education, C-70 37 RFT, U.S. District Court, California (1970).

Hobson v. Hansen, 269 F. Supp 401 (1967).

Larry P. et al. v. Wilson Riles et al., No. C-71 2270 RFP, U.S. District Court, Northern District of California (June 1972).

Marshall v. Georgia, No. CV 482-233, U.S. District Court, Southern District of Georgia (June, 1984).

PASE: Parents in Action on Special Education et al., v. Hannon et al., No. C-74-3586 RFP, U.S. District Court, Northern District of Illinois (July 1980).

Plessy v. Ferguson, 163 U.S. 537, 16 Sup. Ct. 1138, 41 L. Ed. 256 (1896).