AN INVESTIGATION OF MEASUREMENT INVARIANCE IN THE WISC III: EXAMINING A SAMPLE OF REFERRED AFRICAN AMERICAN AND CAUCASIAN STUDENTS

by

CAMETRA L. REED

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Psychological and Social Foundations
College of Education
University of South Florida

August 2000

Co-Major Professor: George Batsche, Ed.D.
Co-Major Professor: John Ferron, Ph.D.

Office of Graduate Studies
University of South Florida
Tampa, Florida

CERTIFICATE OF APPROVAL

Ph.D. Dissertation

This is to certify that the Ph.D. Dissertation of

CAMETRA L. REED

with a major in School Psychology has been approved
for the dissertation requirement on July 11, 2000
for the Doctor of Philosophy degree.

Examining Committee:

Co-Major Professor: George Batsche, Ed.D.

Co-Major Professor: John Ferron, Ph.D.

Member: Kathy Bradley-Klug, Ph.D.

Member: Dierdre Cobb-Roberts, Ph.D.

This document is dedicated first to my Lord and Savior, Jesus Christ. It is He who keeps me, protects me, instructs me, guides me and has comforted me throughout this process. Thank you God for your faithfulness.

Second, to my biological family, I thank you for your love and support throughout my graduate career. Thanks for all of assistance and prayers that you provided. Words cannot adequately express how grateful I am for all that you've done. Continue in God's grace.

Third, to my church family, thank each and every one of you who prayed and keep me before the throne of grace throughout my graduate school experience. I am truly grateful for all of the kind words and encouraging thoughts that have passed my way. Thank you for your financial assistance and physical help in times of need. May God continue to bless you for your hospitality far greater than I ever could.

Fourth, to my university family, thank you to all the faculty, staff and classmates of the University of South Florida that have gone out of their way to insure that I had an insightful and knowledge-filled journey during my graduate school years. I am humbled by the generosity and assistance that I've been afforded during my tenure. I pray that the relationships that have been formed continue to develop professionally and personally.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

AN INVESTIGATION OF MEASUREMENT INVARIANCE IN THE WISC III: EXAMINING A SAMPLE OF REFERRED AFRICAN AMERICAN AND CAUCASIAN STUDENTS

by

CAMETRA L. REED

An Abstract

of a dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Psychological and Social Foundations
College of Education
University of South Florida

August 2000

Co-Major Professor: George Batsche, Ed.D.
Co-Major Professor: John Ferron, Ph.D.

The American educational system has been frequently charged with discriminatory practices regarding the treatment of minority groups. Specifically, African American students have been thought to achieve intellectual and academically below other ethnic groups. The misconception of underachievement led to and was reinforced by systematic discriminatory practices such as ability grouping, tracking and overrepresentation in educable mentally handicapped special education programs. One controversial issue has been the overrepresentation of African American students in the special education process. The roles that teachers, school personnel and school psychologists play, from the referral through the assessment given, are crucial to the inquiry of why African Americans experience differential educational outcomes in the public school environment. To further investigate the trend of overrepresentation, we focused on the intellectual measures given and the presence of construct bias. Specifically, the WISC III was discussed because of it being the most frequently used IQ measure. One emergent technique to assess measurement invariance has been multi-sample confirmatory factor analysis (MCFA). The purpose of this research study was to conduct a multi-sample confirmatory factor analysis of the WISC III to determine measurement invariance between African American and Caucasian students.

Using MCFA, the WISC III scores of 545 African American and Caucasian students in the Hillsborough County Public School System were examined to test the presence of measurement invariance. Multi-sample confirmatory factor analysis provided a more direct comparison in the investigation of factor structure equivalence across groups. A four step series of analyses was conducted during which all possible parameters (factor loadings, the factor correlation, factor variances, and subtest unique and error variances) were constrained for both groups. From the results obtained there were no statistically significant differences in

and error variances) were constrained for both groups. From the results obtained there were no statistically significant differences in the two factor model of th[illegible] the sample of African American and Caucasian students. Which each series of analysis there were no statistically significant changes in chi square or decline in model fit for either group. Therefore, the proposed two factor model as delineated in the WISC III manual provided a relatively good fit to the sample data.

Abstract Approved: _____
Co-Major Professor: George Batsche, Ed.D.
Professor, Department of Psychological & Social Foundations

Date Approved: 7-11-2000

Abstract Approved: _____
Co-Major Professor: John Ferron, Ph.D.
Associate Professor, Department of Educational Measurement

Date Approved: 7-11-2000

Chapter 1

Introduction

> "We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty, and the pursuit of Happiness. That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed....." Declaration of Independence, 1776.

The proclamation of independence of the new Americas from the British monarchy marked the culmination of a political process that had begun in simple protest. The colonists fought against oppressive restrictions on colonial trade, manufacturing, and political liberty. The struggle had developed into a revolution resulting in the establishment of a new nation (Microsoft Corporation, 1998). The new nation began with the task of bringing people from different socioeconomic positions into solidarity. The construction of a unified country proved to be a formidable task, met frequently with sectional loyalties and disagreement. The difficulties were exacerbated in the response of leaders to the institution of slavery. Confusion existed because the institution of slavery was defended in the Southern states as their source of income, while the Northern states felt they were providing the financial resources for the country through industry (Tindall, 1988). The conflict grew into the Civil War. The

Civil War provided an abolitionist platform for the debate in stating that all men were created equal and should be free.

While the Civil War provided slaves with freedom, it did not define the ideology that "all men are created equal." The battle for equality continued for the next 100 years through the Civil Rights movement of the 1960's, the Women's Rights movement of the 1970's, and the Gay Rights movement of the 1980's (Tindall, 1988; Microsoft Corporation, 1998). Each of these movements is in a different phase of existence today. While society struggles with equality, so does the educational community struggle with the ideology of providing equal educational opportunities for Americans.

It has been stated that education does not exist in a vacuum, in fact, it was described as a "microcosm of society" (Shapiro, Benjamin & Hunt, 1995). Therefore, the United States societal ills, woes and struggles are reflected in the public school system. Oakes & Wells (1998) stated that 20th century American society was characterized by financial and power stratification along gender and race lines. With the appearance of an equitable existence, the determination of success verses failure then was based on "meritocracy". Meritocracy was defined as the belief that the race for social rewards was fair, therefore the most successful were more meritorious. In other words, there was a "natural" sorting process (Lewontin, 1992). Nevertheless, the "natural" sorting process has lead educational practices such as tracking and ability grouping (Kelly, 1972; Gamoran, 1992; Kershaw, 1992; Oakes, 1995). Schafer and Olexa (1971) researched the negative effects of tracking on academic achievement. The students in the high tracks received higher academic achievement scores than the

students in the lower tracks. Oakes (1995) found that frequently African American students were underrepresented in higher tracks and overrepresented in lower tracks. Gamoran (1992) found that the process for assigning students to honors-level English classes was biased. However, there are some researchers that believe the tracking system is beneficial in that it is a efficient, effective system that allows students to work at their own pace within their various levels of ability (Kershaw, 1992). Others have found that tracking has cumulative effects that work to maintain the current socioeconomic stratification (Braum, Nelso, & Dykstra, 1975; Franklin & Resnick, 1973; Kershaw, 1992; Webster, 1974). The United States public education system has endured educational reform movements in an attempt to correct these problems.

The reform process has occurred in three distinct phases (Rossmiller, 1987). The goal of phase one was to provide equal opportunity to access education. By the end of the phase, the majority of the population was enrolled in free, public education (Rossmiller, 1987). In 1870 there were 6.8 million pupils in public schools and by 1920, the number had risen to 21.6 million (Tindall, 1988). Phase two was marked by the beginning of the 20th century. One catalyst for this phase was the "space race" towards placing a man upon the moon. When the United States lost this race, the federal government allocated funds and resources to educational programs. Through the National Defense Education Act of 1958, money was funneled into school mathematics and science programs. For minorities, the second phase focused on providing equal opportunity in terms of equal access to education programs with the minimum standards of basic reading and mathematics knowledge. The second phase focused on providing the student with the outcome (reading, writing variables) and neglected the

processes used to achieve them. The third and current phase can be characterized by the focus on equitable educational processes as opposed to final products or outcomes (IDEA, 1997). The reformers are not interested in just ensuring that persons can read, but that they were treated fairly and equally during their educational career.

Also in phase three, schools are questioned regarding the performance of our children in comparison to students in other industrialized nations. Stedman (1994a) reported findings from the 1992 National Assessment of Educational Progress (NAEP) test of seniors in high school. Thirty-seven percent of the seniors tested demonstrated proficiency in reading, the functional illiteracy rate hovered around 20-30 percent of the population, and less than half of those tested appeared to have a firm grasp of 7th grade mathematics concepts. In 1992, the United States ranked 14th out of 15 countries whose population was sampled on the average percentage of math problems solved correctly (National Center for Educational Statistics, 1992). The statistics indicated that 42 percent of 13 year old students in the United States scored at the 50th percentile on international academic achievement tests in language arts, mathematics and science (Stedman, 1995).

Various claims have been made regarding the equality of funding across schools within a school district and school districts within states. Schools receive funding based on property taxes. Many states have attempted to equalize geographical differences through funding formulas, which are not always successful. For example, in 1983 the state of Maryland argued that 100 percent equality in funding was "too expensive" and that the goal would be 75 percent equality (Kozol, 1991). In investigating differences in schools in metropolitan United States cities (e.g., New York, New York; St. Louis, Missouri; Chicago, Illinois; Washington, D.C.), Jonathan Kozol (1991) found that this

type of funding resulted in fewer resources for poor, inner city schools. Inner city schools are populated predominately by minorities, thus affirming segregation, discrimination and differential educational outcomes based on geography and property taxes (Orfield, Bachmeier, James & Eitle, 1997). In the Hillsborough County, Florida school district, allegations have been made that African American students receive an inequitable education based on their school location (Look Beyond the Buses, 1994).

In an attempt to provide "equity" to educational outcomes, the public education system has created a separate special education system to serve students with disabilities. The United States Government implemented legislation in the form of the Individuals with Disabilities Education Act (Dwyer & Stanhope, 1998) to ensure that children with disabilities receive a free, appropriate education in the least restrictive environment. Common issues that face minority children in the public school system included over representation in mentally handicapping special education programming and under representation in gifted special education programming (Dunn, 1968; Smith, 1983; Chinn & Hughes, 1987; Artiles & Trent, 1994). In fact, Patton (1992) suggested that the identification of gifts and talents for African American students is flawed when the assessment strategies used are representative of White Protestant Anglo-Saxon values. The author stated that the relationship among African American worldviews and psychoeducational assessment theory should be investigated to develop a more appropriate assessment strategy.

Minority group students are disproportionally placed in special education categories more frequently than other groups (Barona & Faykus, 1992; Ysseldyke & Algozzine, 1983). For example, Serwatka, Deering, and Grant (1995) found that 58 of 67

Florida school districts overrepresented African American students in classes for the emotionally handicapped (EH). The overrepresentation issue was not confined to the EH classes, but extended to other special education categories. In S-1 v. Turlington (Reschly, 1987) a class action suit was brought against the state of Florida on behalf of African American children receiving special education services under the educable mentally retarded (EMR) category. However, the Judge sided with the State of Florida citing that the plaintiffs did not prove that the African American students had been misclassified in the special education categories (Reschly, 1987).

Court cases against school districts and states concerning the overrepresentation of African American students in mentally handicapped special education classes and the underrepresentation of African American students in learning disabled and gifted programs abound in the 1970's and 1980's with African Americans demanding racial equity in the public schools (Larry P. v. Riles, Marshall et al. v. Georgia, PASE v. Hannon). As a result of the Larry P. v. Riles (1979, 1984, 1986) case, the use of intelligence tests in California was banned because they were declared biased and their use purportedly resulted in disproportionate numbers of African American children being placed in programs for the educable mentally retarded (MacMillan & Balow, 1988). The Larry P. v. Riles 1986 litigation extended the ban to include evaluations for placement in learning disability categories. As a result of the Larry P. v. Riles decisions, another case was filed declaring that it was unconstitutional to ban intelligence testing in all special education categories. Crawford v. Honig (1991) suspended the 1986 ruling that banned intelligence testing in all special education categories and upheld that intelligence tests should not be used in the determination of placement in educable mentally handicapped

categories. The original Larry P. v. Riles verdict is still challenged today however, the attempts have been unsuccessful (CASP v. California Department of Education, 1994). The continued litigation concerning appropriate placement of children has caused the African American community to question America's intentions concerning its children (Hillard, 1973; Jackson, 1985).

The litigation also has resulted in controversies regarding eligibility, identification, and placement practices that are used for minority students who reside in different sociocultural and socioeconomic environments than their Caucasian, middle-class counterparts. School psychologists, through formal training in the assessment of children, play a major role in the determination of the appropriate educational setting for students with disabilities. The 1997 School Psychology: Blueprint for Training and Practice II outlined the responsibility of school psychologists to be

> well versed in a variety of assessment methods, including informal and formal test administration, behavioral assessment, as well as assessment methodologies to define a student's problems and needs, to assess current status, and to measure the effects of a problem-solving process (p. 7).

School psychologists traditionally have been assigned the responsibility of providing psycho-educational assessment batteries to the special education programming process (Macmann & Barnett, 1997). Intelligence testing was made a required part of the psycho-educational assessment protocol in determining special education placement through the Individuals with Disabilities Education legislation (revision and reauthorization of Public Law 94-142) and state regulations (Florida Department of

Education, 1998). It has been estimated that 1-1.8 million intelligence tests are administered to children each year in the United States (Gresham & Witt, 1997; Macmann & Barnett, 1997; Reschly, 1998). Surveys have found that school psychologists spend approximately two-thirds of their time in special education eligibility determination (Curtis, Hunley, Walker & Baker, 1999; Reschly & Wilson, 1995). However, the scores produced by intelligence testing may not contribute to the successful implementation of instructional programming for children. Gresham & Witt (1997) discuss the literature regarding the validity of intelligence tests in providing relevant information in educational planning. The authors assert that the major reason why intelligence tests do not contribute to the process is because there is not sufficient evidence of an aptitude by treatment interaction. By definition, an aptitude by treatment interaction is a belief that the measurement of valid aptitudes (characteristics or traits) predicts the probability of success under certain treatment conditions (educational programs). In the special education decision making process, this would mean that a student's scores on the Wechsler Intelligence Scale for Children- Third Edition (WISC-III) Verbal and Performance indices would be predictive of specific educational contexts in which they would learn. For example, a student with a specific profile of high verbal achievement would learn best through phonics reading approaches. However, the authors cite that the school psychology and special education literature do not support such interpretations (Arter & Jenkins, 1979; Ayres & Cooley, 1986; Ayres, Cooley, & Severson, 1988; Das, 1995; Das, Naglieri, & Kirby, 1995; Good, Vollmer, Creek, Katz, & Chowdhri, 1993; Kavale, 1990; Kavale & Forness, 1987; Ysseldyke & Mirkin, 1982).

Yet school psychologists are bound by law to provide these assessments during the special education decision making process.

As a profession, school psychologists examine current educational practices and the alternatives for more effective education for all students. However, researchers found school psychologists may make decisions that "confirm" the referral question rather than investigating the need for special education services (Huebner, 1989; O'Reilly, Northcraft & Sabers, 1989; Ward, Ward, & Clark, 1990). Frequently, minority students are referred for special education services in higher numbers than the majority culture (Dunn, 1968, 1973; Franks, 1971; Holtzman, 1985). Several authors have documented that bias exists in the decision making process for special education (Fuchs, 1987; Grossman & Franklin, 1988; Reilly, 1991). In summary, these studies have found that disproportionally, minority students are referred for special education services and are usually given a special education categorization and receive special education services.

Additionally, with the viewpoints of some such as Jensen (1994) that African American children's school performance differs because of genetic origin further strengthens the position of many African Americans who believe that the majority United States culture supports an "inferior" minority argument (Herrnstein & Murray, 1994; Levin, 1994; Peoples, 1995). There is a strong position that cognitive ability score differences exist because of a lack of ability of minorities (Reynolds & Jensen, 1983). However, researchers also have not found statistically significant racial differences developmentally and with academic achievement (Entwisle & Alexander, 1988; Rowe, Vazsonyi, & Flannery, 1994). The position that there are ethnic differences in cognitive ability created hostility and mistrust in minority groups regarding the proper use of

intelligence testing and raise issues of test bias or construct validity (Gould, 1981; Medina & Neill, 1990).

Construct validity attempts to assess whether intelligence is actually represented by test performance on a specific measure (Hopkins, Stanley & Hopkins, 1990). The Standards for Educational and Psychological Testing (1985) state that "the evidence of construct validity ...for a particular test should be embedded in a conceptual framework...." (p. 9). Messick (1989) suggested that construct validity is the overarching form of validity. A test that has construct validity can demonstrate that it measures what it is supposed to measure across various ethnic minority groups (Anastasi, 1988). If a psychological test measures different hypothetical traits (intelligence) for one group than it does for another group then the test may have construct bias (Keith & Reynolds, 1993). Therefore the proposed study will provide information in the investigation of the construct bias issue of a popularly used psychological measure of intelligence or "g", the Wechsler Intelligence Scale for Children- Third Edition (WISC-III).

Ethnic group differences and bias have been measured in the Wechsler Intelligence Test series over the past twenty years. There are studies that substantiate the claim that there are differences in Full Scale IQ scores between racial groups (Kaufman & Dolppelt, 1976; Reynolds & Gutkin, 1980) and that these scores remain stable over time (Elliot & Boeve, 1987). However when investigating the subscale scores, some studies have found African Americans perform better in the Verbal subscale than in the Performance index. Vance, Hankins & McGee (1979) studied students who had earned low IQ scores on the WISC-R. The authors matched samples of African American and Caucasian males and females based on gender, chronological age and Full Scale IQ

scores. The researchers found that the African American sample obtained higher Verbal scores than the Caucasian population. Munford (1978) found similar results when investigating African American psychiatric outpatients. While these studies provide support for different IQ score patterns between ethnic groups, the authors used previously referred and placed samples of African American students. However Taylor, Ziegler, & Partenio (1984) conducted research using randomly selected samples stratified for age, gender and ethnicity. The authors found statistically significant larger Verbal-Performance differences in Hispanic students than in the African American and Caucasian students. Additionally, there were significant differences found between the overall IQ scores obtained by the Caucasian (112.36), Hispanic (99.39) and African American populations (96.17). Reynolds (1980) researched the construct validity of the WISC-R through investigations of ethnic raw score differences over time using the standardization sample data. His results indicated that African American males showed the smallest amount of change in raw scores across time. Thus, Reynolds concluded, providing evidence that African American "mentally" develop at slower rates than their Caucasian counterparts. With the continued differences found in Full Scale IQ scores, additional research has been conducted to provide validity evidence in intelligence testing using the WISC series.

One of the common methods of examining validity in intelligence tests is factor analysis. The factor analysis process allows the researcher to examine differences in groups on the test items. It may be used to investigate group differences (exploratory) or confirm existing models (confirmatory). Valencia, Rankin, & Oakland (1997) found that the factor structure of the WISC-R was not supported for ethnic minority groups when

using more stringent statistical analysis methods (LISREL8). When using exploratory factor analysis, the 2nd (Perceptual Organization) and 3rd (Freedom from Distractability) factors were in reverse for the African American and Mexican American students in the sample. Additionally, when using confirmatory factor analysis methods the authors found most of the variance on a general "g" factor as opposed to additional factor models. In summary, the Valencia, Rankin & Oakland (1997) study found different factor structures in the WISC-R with ethnic minority samples.

Slate and Jones (1995) used confirmatory factor analysis to examine the validity of the WISC-III with African American students undergoing special education evaluations. The authors found that the initial unrotated factor structure of the WISC-III tended to remain the same for this population of students as it had been identified in the test manual. However, Slate & Jones found that when rotating the factors (making the model more parsimonious), the Arithmetic, Digit Span and Coding subtests did not load on the factors that had been identified in the test manual. Thereby providing evidence that even though students may have similar profile scores on the Verbal scales, there are differences in the numerical abilities of African American and Caucasian students as measured by the WISC-III. From the studies discussed thus far there is evidence that differences exist between various ethnic groups as measured by the WISC III.

In summary, African American and Caucasian students have been shown to perform differently on this test. However, substantial evidence does not exist to support the reasons for these differences. The researchers often used African American samples that had been identified to receive special education services or were in the evaluation process. Some of these studies employed the most stringent statistical methods of that

time in using exploratory and confirmatory factor analysis. The present study will extend the literature in providing evidence of the construct validity of the WISC-III for use with an African American population by examining a sample of African American and Caucasian students in Hillsborough County, Florida using multi-sample confirmatory factor analysis.

Purpose

The purpose of the current study is to investigate the Weschler Intelligence Scale for Children-Third Edition (WISC III) as a measure of intelligence for African American children. Using the multi-sample confirmatory factor analysis method, the researcher will examine the degree to which the WISC-III measures the same underlying constructs for a sample of African American students at it did for the majority group. This investigation will analyze the factor structures of the WISC-III for samples of African American and Caucasian students using the data obtained from the Hillsborough County Public School System database of students referred for special education evaluation.

Research Questions

1. Is construct bias present or absent in the WISC-III when comparing a sample of African American and Caucasian students?

2. Are the factor loadings invariant across the African American and Caucasian student populations?

3. Are the factor variances and the factor correlation invariant across the African American and Caucasian student populations?

4. Are the subtest unique and error variances invariant across the African American and Caucasian student populations?

5. Are the covariance matrices of subtest scores equivalent for African American and Caucasian students?

Hypotheses

1. There will be no statistically significant difference when comparing construct bias in the WISC III for a sample of African American and Caucasian students.

2. There will be no statistically significant difference in the factor loadings of the WISC III when comparing a sample of African American and Caucasian students.

3. There will be no statistically significant difference in the factor variances and the factor correlation of the WISC III when comparing a sample of African American and Caucasian students.

4. There will be no statistically significant difference in the subtest unique and error variances of the WISC III when comparing a sample of African American and Caucasian students.

5. There will be no statistically significant difference in the covariance matrices of subtest scores on the WISC III for a sample of African American and Caucasian students.

Chapter 2

Review of the Literature

The literature review first will focus first on the history and trends in the education of African American students in the United States. Within this context, first the discussion begins with a review of some of the popularly held beliefs regarding African American students' abilities in intelligence and academic achievement. It has been suggested by some that African Americans obtain lower scores on intellectual and achievement measures because of inherent cognitive deficits (Murray & Herrnstein, 1994; Reynolds, 1980). Second, the practice of tracking and its contribution to educational differences will be discussed. Third, the discussion will focus on the literature on how African American students are affected by the inability of the regular public education system to meet their needs. The public school system's overrepresentation of African American students in certain special education categories will be discussed. This section will be concluded with the discussion focusing on 20 years of litigation on the misrepresentation of African American children in the special and regular education systems. The African American community increasingly has become aware of the controversies and inadequacies of the public education system and has responded with legal action. Frequently, African American students are placed in classes for the mentally handicapped at higher rates than their Caucasian counterparts.

Because of the overrepresentation of African Americans in certain special education categories, it is necessary to examine the process in which these placement decisions are made. The literature review will investigate the problem of bias in the special education placement process. First, the research regarding school personnel bias in referrals will be discussed. Second, the role of the school psychologist in special education placement will be highlighted. Because the school psychologists' evaluation is based on various tests, the literature review will conclude with a discussion of the relationship between the special education placement process and evaluation assessment. First, the history of the most frequently used intelligence tests, the Wechsler series, will be discussed in relation to construct validity. Specifically, the discussion will focus on the investigation of ethnic differences in the Wechsler tests. Finally, the literature review will focus on the use of multi-sample confirmatory factor analysis methods as an approach to investigate construct validity.

History and Trends in the Education of African American Students in the United States

African Americans, intelligence and academic achievement. During the development of the independent African American community in the United States, there has been a persistent debate regarding the intellectual capabilities of African Americans in comparison to Caucasians (Walters, 1995). Smedley (1993) outlined the development of ideology regarding the biological basis for race and exemplified the differences between the races. The historical anthropology literature provided contradictory evidence for the presence of race differences based on physiological traits and patterns. The author concluded that the question of racial differences was a socio-cultural phenomenon as opposed to a biological fact. Much of the differences quoted in research and position

papers were based on social stratification and the desire the keep the Negro at the lowest social rung (Smedley, 1993). In the 1950's, during the battle for desegregation in the public school system the question of equality of intellectual capabilities was raised. Some suggested that there were biological differences between the races that substantiated the belief that African Americans were not able to perform equally on intelligence tests (McCurk, 1951). Karier (1972) suggested that the roots of the testing movement began with the need for workers during World War I. Many of the intelligence tests that were developed were based on the presumption that there was a hierarchical occupational structure. Meaning, social class was a fair predictor of intelligence. Therefore, most African Americans and other minorities who fell at the lower ends of the social class were less intelligent and delegated to menial or blue-collar jobs and that their was a biological basis for this phenomena.

However, a statement was submitted to the Supreme Court by 30 American social scientists suggested that the variations between ethnic groups in intelligence measures may have been a result of environmental differences and were not "biological" (U.S. News & World Report, 1956). The debate over the intellectual capabilities of African Americans continued throughout the Civil Rights era of the 1960's and 1970's with support from both sides. In 1971, Herrnstein published an article that suggested that heredity explained 80-85% of the variance in intelligence. To such arguments rebuttal quickly came in stating that there are no data in existence that support the hypothesis that IQ test scores are heritable (Kamin, 1974). The debate has been heightened in the 1990s with the publication of the book, *The Bell Curve* (Murray & Herrnstein, 1994). The researchers proposed that the differences in intellectual and academic attainment between

African American and Caucasian students are due to genetic differences in ethnic groups. Again, rebuttal has come that the differences in IQ scores may be due to socio-cultural factors, socioeconomic factors, a lack of cultural responsiveness of the public school system, or insensitivity to cultural differences (Russo & Talbert-Johnson, 1997; Villegas & Watts, 1991; Wagner, 1995).

There are racial stereotypes that support a belief that African American students are intellectually incapable of achieving on the same level as the majority culture. However, research evidence exists that provides information on the commensurate developmental progress between racial groups (Rowe, Vazsonyi, & Flannery, 1994). Although all children do develop similarly regardless of ethnicity, there is still a difference in the intellectual and achievement scores achieved on standardized assessment instruments between ethnic groups (National Center for Education Statistics, 1992). From the research evidence presented thus far, it is clear that African American students are challenged with culturally and racially insensitive systematic educational practices when they enter the classroom. And because of compulsory education laws within America, they must still enter the classroom. The next question becomes, "How do they perform once they are there?"

Numerous studies (e.g., Herrnstein & Murray, 1994; Levin, 1994; Peoples, 1995) have been conducted that analyze the differences in achievement scores among the different races. The conclusions drawn from these data lead many to erroneously conclude that minority groups are just not as smart as the White majority. This literature review presents research that dispels the myth that achievement differences are based on

intelligence, by showing that many other factors contribute to high academic achievement.

The current stereotypes of African American student performance in the classroom are upheld by research that purports differences in the intelligence of the two groups. One such study was conducted by Reynolds and Jensen (1983) using a popular intelligence test, the Wechsler Intelligence Scale for Children- Revised (WISC-R). Reynolds and Jensen compared 270 Caucasian and 270 African American children drawn from a national, stratified random sample used in the WISC-R's standardization data. The students were matched on sex, age, and WISC-R Full Scale IQ scores to examine subtest differences between the two racial groups. The authors conducted a multivariate analysis of variance of group differences across all subtests. Tests of significance also were conducted to examine group differences on the three main factor scales (Verbal, Performance and Memory), which contribute to the variance within the overall test. The results suggested that there were no significant differences between the groups on the Verbal scale. On the Memory scale, the African American students' scores exceeded the Caucasian students' scores. Finally, on the Performance scale the Caucasian students' scores exceeded the African American students' scores. When the authors "matched" the two samples on demographic and not cognitive variables, they found that the African American sample obtained higher mean scores in only one subtest (Digit Span) of the Memory scale. While the authors noted that these differences were equivalent to less than three Full Scale IQ points, from these results, the authors concluded that "black-white differences are due primarily, but not entirely, to differences in general ability" (p. 213).

Although the authors conducted an empirical study and found statistically significant results, the inflammatory conclusions drawn tend to support racial stereotypes that African American children are less intelligent than their Caucasian peers. The Performance scale mean difference between the African American and Caucasian samples was .176. The Memory scale mean difference also was .176 for the African American and Caucasian samples. However, the standard deviation was on the average .822 for the Performance scale and on the average .748 on the Memory scale. Thus, the results cited lack clinical significance. Meaning that the results found would not have practical significance in interpreting results for an individual student. A critical analysis of the results shows that although statistically significant, overall average mean differences of .129 between the groups are not clinically significant with a standard deviation of .857. Overall, the differences found between the groups were not large enough to support a conclusion that African American and White children are different in general ability.

A study by Rowe, Vazsonyi, and Flannery (1994) also examined claims made about differences between races. By investigating the developmental processes of White and African American children, the authors concluded that developmental processes were nearly identical across races. Developmental processes referred to the association among specific variables within each ethnic group (e.g., family structure, academic achievement) to the total variance of the identified variable across all groups. Using data sets from seven studies which included over 7,000 youths from across the United States, the authors compared covariance matrices using the linear structural equation modeling to determine group differences (Rowe, Vazsonyi, & Flannery, 1994). The variables

included in the matrices reflected factors such as family and home characteristics, peer and environmental influences, and academic adjustment. Overall, the authors found indistinguishable developmental process differences between racial groups.

However, there are threats to the external validity of these findings. In this study, the authors completed a meta analysis of various studies. In conducting a meta analysis, a caution must be taken to avoid a Type I error, meaning that the effects found may not be true effects. The effects found may be attributed to the combination of numerous studies with large effect sizes. Despite this limitation, the lack of developmental differences encourages additional research into other variables that could influence academic achievement differences between races.

A study by Entwisle and Alexander (1988) analyzed the achievement differences found between African American and White students not only on published standardized achievement instruments, but also on classroom grades. Using a stratified random sampling procedure, 307 African American and 275 White first grade students were selected based on their promotion at the end of the first grade. The authors collected data through parent and child interviews, teacher ratings, cumulative records reviews and reviews of test scores on the California Achievement Test (CAT). The two racial groups were compared on the following variables: race, sex, parent's educational attainment, parent's general ability estimate, parent's expectations, personal maturity, special problems, expectations of their self, grades, peer popularity, absences, and CAT scores. These data were collected at four different times during the year, Fall, first quarter, third quarter and Spring. Scores on these variables became factor determinants in contributing to students' grades and test scores.

The authors did not specify their method of data analysis. The results showed that although the children were equal on the achievement and parental variables at the beginning of the year, racial differences began to appear after the first grading period. Although background characteristics such as higher scores on the CAT and higher parental expectations would lead to predictions that the African American student's initial grades would be higher, the opposite was found. A small (.20) but critical raw score difference was found between the two racial groups, favoring White students. Over time, the number of factors that contributed to the grades and test scores of African American children decreased while remaining the same for White children. Therefore, African American children had fewer things they could do to improve their grades.

The relation between initial grades and race were found to have deleterious effects on African American student's academic progress. The initial low grades were predictors of later grades, which were even lower, thus leading to a gulf in the achievement scores of African American and White students. The differences in school grades also were predictive of the scores on the CAT. What began as a statistically insignificant three-point difference in scores between the races in the Fall developed into a statistically significant eight-point difference by the Spring. The authors concluded that race is a significant factor in the grading of elementary students which negatively impacts the achievement of African American students.

In conclusion, the African American population's differential scores on standardized achievement measures and classroom grades baffle researchers. Although the students in the samples were matched on the majority of characteristics at the outset, White students still maintained higher achievement scores than African American

students in the end. Although areas were found which theoretically should have lead to higher rates of achievement for African Americans, in reality, these areas did not significantly impact the African American student's scores. Results such as these lead to more research questions concerning the different academic growth patterns of the two racial groups.

Tracking. The United States education system has misrepresented the interests of African American children by using homogeneous ability grouping or tracking as a mechanism for education. The disturbing aspect of tracking is that minority students tend to be overrepresented in the lower academic tiers (Oakes, 1986; Villegas & Watts, 1991). Karier (1972) gave Thorndike's justification for the implementation of the tracking system. Thorndike, the father of the education tracking, described the initial tracking system as a way to "classify and standardize the school's curriculum with a differentiated track system based on ability and values of the corporate liberal society" (p. 247). However, the author noted that researchers have found that this leads to a self-fulfilling prophecy when a child was objectified or tracking. When compared to the "higher ability" peers, "lower ability" students are called on less often in class, given less time to respond, praised less frequently, given less feedback, criticized more frequently, and prompted less often in the case of incorrect responses (Cazden & Mehan, 1989; Good, 1970; Hilliard, 1989; Irvine, 1990; Lehr & Harris, 1988; Rist, 1970).

Twenty-eight years ago, a classic study conducted by Schafer and Olexa (1971) demonstrates the negative effects of tracking on the achievement of African American students. The academic achievement scores of 1,157 students from two mid-Western high schools were evaluated to test the effects of college preparatory versus non-college

preparatory tracks for students. The racial distributions of the two schools were: 77% White and 23% African American for one and 92% White and 6% African American for the other. Of those students combined, 72% of the White students and 30% of the African American students were placed in the college preparatory track, while 28% of White students and 70% of African American students were placed in the non- college preparatory track. By analyzing the weighted percentages, the authors found that 66% of the college preparatory track students received high and high average achievement scores while 82% of the non- college predatory track students received low average to low achievement scores as defined by grade point average. The authors also evaluated the amount of academic change over the high school years. The tendency was for college preparatory tracked students to make increasingly higher grades, while the non-college preparatory students made increasingly worse or stabilized grades. Rarely did students move between tracks. Throughout the study, only seven percent of the student population made a transition between either of the tracks. Schafer and Olexa also found that students in the non-college preparatory track experienced less participation in extracurricular activities, greater misconduct in school, a greater tendency to drop out, and greater delinquency.

One of the limitations of the Schafer and Olexa study was that the geographic boundaries of the study prohibit generalization to other areas of the country. Therefore, the reader is cautioned in drawing inferences about their data to any other school (Shafer & Olexa, 1971). Although this caution is warranted, it is suspected that the effects may be very different in more heterogeneous communities because of the diversification of the general population. The authors also listed a number of confounding or extraneous

variables such as peer pressure and parental influence that also could have affected the student's scores. However, they selected a sample that was representative of all IQ scores and family income variables for the population as an attempt to correct for family, peer and community variables which could affect achievement scores. Finally, the authors noted that the study measured only end variables and "did not permit the determination of what psychological, interpersonal, or contextual factors intervened between track position and the educational and behavioral outcomes" (p. 32). Hence, the data obtained provided descriptive information and focused greater attention on the looming question of why track position makes a difference.

The preceding evidence also is supported by a study by Oakes (1995) which examined the tracking systems of two school systems, Rockford Public Schools in Rockford, Illinois and San José Unified School District in San José, California. Data were gathered from student enrollment, achievement, and curriculum documents to examine the impact of tracking on student outcomes. Oakes analyzed these data to report in litigation cases against both school districts concerning improper racial tracking. The method of analysis consisted of content analysis of the curriculum documents, descriptive statistics comparing the achievement ranges of different tracks, and probability estimates of student placement based on prior achievement.

Oakes found that overall, African American students consistently were underrepresented in higher tracks while White students consistently were overrepresented. The converse was found in the lower tracks; African Americans were overrepresented while Whites were underrepresented. The author attributed these differences to a lack of consistency in applying placement criteria to tracking decisions.

The subjective opinion of the previous year's teacher was the most heavily weighted factor in placement decisions. Placements also were found to be skewed racially in both areas. African American students who received the same standardized achievement score were less than half as likely to be placed in an accelerated track than White students.

The author presented an example where none of the African American students who scored in the top quartile (75-99th percentile nationally) on the California Assessment Program (CAP) reading comprehension test were placed in the high track English class whereas 40% of the White students in that quartile were placed in the accelerated class. Oakes' discussion of archival information for two schools districts did not include any type of statistical analysis which limits the reproducibility of the results. Therefore, the results of her analysis are descriptive of tracking systems and provide supportive qualitative information, but cannot be generalized to other school districts.

Gamoran (1992) found further evidence of minority status being a persuasive basis for educational placement in a study of 1,102 students from five communities representing one suburban, one rural and two urban public school districts in the Midwest. The purpose of his investigation was to examine the processes and criteria of assignment to honors high school English classes. By analyzing data collected from school records and questionnaires, Gamoran used a logistic regression model to estimate the likelihood of assignment to honors English class as a function of ascribed, achieved and organizational criteria. Ascribed characteristics included gender, ethnicity and socioeconomic status. Achieved characteristics included reading, math and writing scores on standardized achievement tests as well as prior grades in English. The organizational criterion was placement in 8th grade English high ability groups. Within the logistic

regression model, three equations were estimated: "one with only the ascribed characteristics, a second which added the achieved variables, a third which takes prior grouping into account,..." (p. 196).

Logistic regression coefficients were reported for the three general model equations. Just considering the ascribed characteristics of gender, ethnicity and socioeconomic status in the assignment to 9th grade honors English classes, a statistically significant coefficient was found for minority status (b= -1.291). When achievement variables, such as standardized test scores, and previous grades in English were taken into account along with the ascribed characteristics, again a statistically significant negative coefficient was found (b= -.794). For the third equation, prior grouping in a high ability English class was considered along with ascribed and achieved characteristics and again, the author found a statistically significant negative value (b= -.976). Gamoran (1992) concluded that "the transition to high school is a critical point for the trajectory of minority students" (p. 202). Even when numerous other variables were taken into account like prior achievement in English, minority students more often were not assigned to honors English classes than any other group.

Gamoran's findings are critical in the discussion on the misrepresentation of African American students in honors' classes. While holding other factors such as previous educational placements, socioeconomic status and scores on standardized achievement tests constant, Gamoran still found evidence that ethnic background was a significant predictor of student placements. Gamoran used schools representing rural, urban and suburban communities increasing the ability to generalize his results to other schools. However, one limitation of the study was that the schools were all located in the

Midwest, limiting the generalization of the results for other geographic regions in the country. In conclusion, Gamoran's study provides further data that minority status is a negative contributor to the tracking or educational placements of minority students.

There are arguments for the validity of tracking. It has been said that one "advantage" of tracking is that it is a positive, viable system that assists in the efficient, effective learning of all students and allows students to work at their own pace within their ability levels (Kershaw, 1992). However, through a critical, historical analysis, Kershaw found that tracking actually acts to maintain the status quo through a number of factors. First, Kershaw reviewed research that found that teachers enter the classroom with expectations about student abilities (Braum, Neilson, & Dykstra, 1975). Kelly (1972) also found that teachers perceive students in lower tracks as possessing less academic ability as other students. With this in mind, numerous investigators have found that students tend to live up to the expectations of their teachers (Cooper, 1979; Dusek, 1975; Eder, 1981; Hargreaves, 1967; Luge & Luge, 1978; Ravitz, 1963; Rosenthal & Jacobson, 1968). Ultimately, with the knowledge that students from lower classes and minority groups generally are placed into the lower track classes (Edelman & Howe, 1985; Kelly, 1972) the conclusion can be drawn that low teacher expectations support the lack of academic achievement of the lower tracked students. This phenomena reinforces the placements of minority students and maintains the status quo.

Finally, Kershaw believes the effects of tracking become cumulative. If at the beginning of a child's education career, teachers place them in lower tracks, then the chain rarely is broken. Records of past academic achievement serve to predict future placement, thereby locking lower track students in a career of underachievement. Schafer

and Olexa (1971) found that the drop out rate among lower track students is higher than students placed in other tracks. Dropping out leads to negative effects on potential financial and vocational success (Franklin & Resnick, 1973; Webster, 1974). High school dropouts earn significantly less than high school graduates and less than one half the income of college graduates (Center for the Study of Social Policy, 1996). The Center for the Study of Social Policy (1996) also found that high school dropouts have a higher likelihood of slipping into poverty. Between 1991 and 1992, 5.4% of high school dropouts became poor while only 2% of high school graduates fell into poverty. Citing 1983 U. S. Bureau of Labor Statistics, Kershaw (1992) found that over 60% of the managers and professionals in society have had at least four years of college. If the lower tracked students (predominately minorities and lower socioeconomic classes) do not have the opportunity to take higher level classes then they likely will not have opportunities to compete for managerial and professional jobs.

In summary, the various studies on the tracking of minority students in public schools indicate that students do not benefit academically from being segregated into different educational tracks. In fact, tracking often leads to poor academic outcomes. Compounding this problem is the fact that the effects of tracking students become cumulative over time. Lower tracked students have been found to receive few, if any, opportunities to move into higher tracks. Being, in a sense, "stuck" in these lower tracks maintains their lower achievement scores on standardized assessment instruments. When students are "stuck" in these lower tracks, their opportunity for success in life is influenced negatively (Center for the Study of Social Policy, 1996). Thus it appears that

one end result of tracking has been to maintain the status quo of racial segregation and social immobility in the United States.

The overrepresentation of African American students in special education categories. More African American students proportionally are placed in special education programs than any other racial or ethnic group (Reschly, 1991). African American students also are underrepresented in programs for the gifted. The majority of system-level educational practices such as tracking or ability grouping and special education services usually have a greater representation of African American students filling the lower tracks or groups. Ultimately, these practices severely impact the achievement levels of these students.

Public schooling historically has struggled with the disproportionate placement of African American children in the special education system. In 1968, Lloyd Dunn called attention to the overrespresentation of minority students placed in classrooms designated for students with mental retardation. According to his "best judgment," 60 to 80% of students taught by special education teachers were minority children with low socioeconomic backgrounds (Dunn, 1968). Dunn's claims were especially pertinent during this time because of the concern for the plight of disadvantaged children that was evidenced through major events such as the Civil Rights movement, the War on Poverty initiative, and the Coleman report (Artiles & Trent, 1994). Also during this time, the Office of Civil Rights began to collect data on the representation of minority students in the special education population. Smith (1983) found that the Office of Civil Rights (OCR) surveys during the 1974-78 period data concluded that African American

participation in classes for the educable mentally retarded was 3.4 times greater than that of whites.

Chinn & Hughes (1987) analyzed data obtained through the Office of Civil Rights Elementary and Secondary Schools Civil Rights survey administered from 1978 to 1984 on minority representation in special education. In 1978, African Americans made up 38% of students served in programs under the administrative category of "educable mentally retarded" (EMR), 27% of students served in programs under the administrative category of "trainable mentally retarded" (TMR), and 24% of the students served under the administrative category of "seriously emotionally disturbed" (SED) in special education populations while only representing 15.7% of the total school enrollment. In 1980, African Americans enrollment in special education categories increased as their total school enrollment increased, however they were still overrepresented in special education categories. African Americans made up 45% of the students served under the EMR category, 30% of the students served under the TMR category, and 28% of the students served under the SED category in special education populations while representing 20% of the total school enrollment. In 1982, the upward trend continued as African Americans made up 54% of the students served under the EMR category, 37% of the students served under the TMR category, and 32% of the students served under the SED category in special education populations while only representing 25% of the total school enrollment. Finally, in 1984 the total school enrollment of African American students stabilized, however, the trend of increasing overrepresentation in special education populations continued. African Americans made up 48% of the students served under the EMR category, 33% of the students served under the TMR category, and 30%

of the students served under the SED category in special education populations while representing 24% of the total school enrollment (Chinn & Hughes, 1987).

The disproportionate representation continued into the 1990s. A study by Serwatka, Deering & Grant (1995) found that in the state of Florida, African American students were represented disproportionately in the special education administrative category of "emotionally handicapped" (EH). Using data collected by the Florida Department of Education, the researchers established a negative Pearson product-moment correlation between the percentage of the student population that was African American and the overrepresentation of African American students in EH classrooms of -.26 (r = -.26, $p<.05$). These results indicate that 58 of the 67 Florida school districts overrepresented African American students in EH classes (Serwatka, Deering & Grant, 1995).

McMillan & Reschly (1998) examined the Office of Civil Rights survey data gathered concerning the overrepresentation of minority students in the mild mental retardation (MMR), specific learning disability (SLD) and seriously emotionally disturbed (SED) special education categories over three decades (1978, 1986 and 1990). The results indicated that there has increase in the participation of Caucasian, African American and Hispanic ethnic groups in special education. The increase can be attributed to more students being classified into the SLD program. For example, the Caucasian student enrollment in SLD was 2.32% in 1978, 4.29% in 1986 and 4.97% in 1990. The African American student enrollment in SLD was 2.23% in 1978, 4.43% in 1986 and 4.97% in 1990. The Hispanic student enrollment in SLD was 2.58% in 1978, 4.31% in 1986 and 4.68% in 1990. In the MMR category, it was noted that the African American

student enrollment had decreased from 1978 (3.46%) to 1986 (2.30%) to 1990 (2.10%). However, the percentage of African Americans enrolled during each of these years was still overrepresented in comparison to the Caucasian (1978-.97%; 1986-.87%; 1990-.81%) and Hispanic (1978-.98%; 1986-.56%; 1990-.65%) populations.

The 1998 Annual Report to Congress, the U. S. Department of Education Office of Civil Rights reported that discrepancies in disability prevalence and service provisions across ethnic categories was most prevalent in the mental retardation category (Office of Civil Rights, 1998). A total of 2.6% of African American students were identified as having mental retardation. In contrast, 1.2% of Caucasian students and .9% of Hispanic students were identified in the mental retardation category. However, African American, Caucasian and Hispanic students were equally likely to receive services under the specific learning disabilities category.

While some progress has been made nationally to decrease the disproportionate numbers of African American students in specific categories, there are still systemic barriers to educational equity present. In the 1998 Report to Congress it also was reported that one state administered standardized tests to determine which students should be placed in classes for the educable mentally handicapped and which should be placed in the classes for the learning disabled. In general, most educators believe that it is more restrictive to be placed in the educable mentally handicapped classes. Despite similar scores of the African American and Caucasian students on the test, nearly all of the African American students were grouped in the educable mentally handicapped classes while the majority of the Caucasian students were placed in the program for students with learning disabilities (Office of Civil Rights, 1998).

Reschly (1987) has offered a warning in the interpretation of overrepresentation data. He suggested that the distinction be made between the percentage of minorities in the total school population, the percentage of minority students that are in special education and the percentage of special education students who are minorities (Reschly, 1987). As an example, Reschly discussed the *Larry P. v. Riles* (1979, 1984, 1986) case in which ten percent of the student population in California were African American. Twenty-five percent of the students placed in classes for individuals with mental retardation were African American. However, Reschly contends, the majority of advocates mistakenly conclude that 25% of African American students were placed in classes for individuals with mental retardation. After applying his criteria, the "correct" conclusion would be that 1% of African American students were placed that year (Reschly, 1987).

Artiles and Trent (1994) argue that the proportion of minority students in the general student population is an important consideration that is neglected by Reschly. For example, Harry (1992a) found a positive correlation between the number of minority students in the general school population and the representation of minority students in special education classes. Additionally, Heller, Holtzman & Messick (1982) found that as the educational program sizes were increased, the more disproportionate the representation of minority students became. Therefore, factors other than methodological weaknesses impact the overrepresentation issue.

Litigation. African American parents increasingly have become aware of the deficiencies of the public school system in meeting the needs of their children. To correct these wrongs, the community has sought litigation in the overrepresentation of

African American children in special education programs. In the past 20 years, four major cases concerning minority representation have passed through the court system with very different consequences.

Plaintiffs in the widely recognized Larry P. v. Riles (1972, 1979, 1984, 1986) case claimed that African American students were overrepresentated in classrooms for students with mental retardation (Reschly, 1991). In 1971, the plaintiffs, African American students in San Francisco, alleged that the biases of intelligence tests were the cause of disproportionate placements of African American students in classes for students with educable mental retardation. The plaintiffs charged that the misclassifications violated the equal protection clause of the U.S. Constitution (Bersoff & Hofer, 1990). As support to their claim, the children were retested by African American school psychologists using culturally relevant language, rapport building, distraction reduction and a change of scoring procedures to encompass theoretically correct answers. The results were that the students scored above the educable mentally handicapped (EMR) range. The defendant, the San Francisco school district responded that the tests were culturally and racially insensitive, but that a better system was not available (Bersoff & Hofer, 1990). Judge Peckham decided in favor of the plaintiffs after citing the biases of standardized intelligence tests and the disproportionate numbers of African American students placed in EMR classes. The final decision in 1986 resulted in a permanent ban of intelligence test usage in the placement processes for African American children in California.

However in 1991, the ban on using intelligence tests for placement in special education programs was challenged in Crawford v. Honig. The plaintiffs claimed that the

prohibition of the use of intelligence tests was a violation of their due process to receive an intellectual assessment as a part of the special education evaluation for specific learning disabilities. Judge Peckham agreed with the plaintiffs and rescinded the ban in using intelligence tests for all special education placements. Yet, the Judge did uphold the ban against intelligence tests when conducting an assessment for mentally handicapped categories. The original Larry P. v. Riles verdict is still challenged today, albeit the attempts to reverse the decision have been unsuccessful (California Association of School Psychologists (CASP) v. California Department of Education, 1994).

Controversial litigation in Chicago (PASE v. Hannon), Georgia (Marshall v. Georgia) and Florida (S-1 v. Turlington) have sought and found different outcomes than were found in the Larry P. decision. In PASE v. Hannon (1980), Judge Grady ruled in favor of the school district that the use of intelligence tests did not violate federal statutes or the U.S. Constitution. Judge Grady personally reviewed test items from the Wechsler Intelligence Test for Children- Revised (WISC-R) and the Stanford-Binet Intelligence Test and found that the complaints had face validity, however, there were only a few questionable items that would not affect a child's overall intelligence score (Bersoff & Hofer, 1990). The Judge suggested that if the psychologists ask questions in a culturally sensitive and intelligent manner, the overidentification problem would be rectified. The plaintiffs attempts for an appeal to the decision were retracted after the Chicago Public School System eliminated the use of intelligence tests in identifying students for placement in EMR (Bersoff & Hofer, 1990).

In Marshall v. Georgia (1984, 1985), a class action suit was filed on behalf of all African American students in special education in the state of Georgia claiming

violations of civil and constitutional rights including the equal protection clause of the 14th Amendment and the Civil Rights Act of 1964 (Reschly, Kicklighter, & McKee, 1988). The plaintiffs alleged that overrepresentation was caused by: (a) procedural violations such as timeliness of reevaluations, (b) individual education plan development and review, and (c) improper interpretation of the federal and state requirements governing classification such as intelligence test cutoff scores (Reschly et al., 1988). The defendants, the state of Georgia, avoided the controversy over the applicability of intelligence tests to minorities and focused on the positive benefits associated with programs for the educable mentally retarded. The judge ruled in favor of the defendants because of the lack of substantial information proving that misclassification was evident provided by the plaintiffs. One unique element of this decision was that the court did agree with the plaintiffs that an overrepresentation of African American students did exist in the EMR special education category (Reschly et al., 1988).

Finally, in Florida, a class action suit was brought against the Florida State Department of Education on behalf of African American children receiving EMR special education services alleging that numerous Education of All Handicapped Children Act (1975) procedures had been violated (Reschly et al., 1988). The plaintiffs claimed that the intelligence tests used were biased, the practices of the school districts were discriminatory and that the least restrictive environment clause had been violated. The defendants rebutted with the same strategies used in the Marshall case, the use of a multimodal identification approach for classification and a strong emphasis on the benefits of EMR programs to all students with low intellectual functioning and learning problems (Reschly et al., 1988). Judge Atkins sided with the defendants in that the

plaintiffs failed to prove that African American students had been misclassified. Therefore, in each of these cases, the courts ruled that the plaintiffs did not provide adequate documentation that the overrepresentation of African American students in these special education programs was detrimental to their educational achievement.

In conclusion, the legal ramifications of overrepresentation of African American students in special education courses have resulted in more controversy. Because of the differences in the Larry P. and Marshall decisions, the courts have allowed the necessity of African American student placement in special education alternatives to be left open for interpretation. Although African American students are placed in special education classrooms at a higher rate than other minorities, litigation spurred from this injustice insufficiently protects the equity of education for African Americans.

Bias in the Special Education Placement Process

In examining the concern regarding minority overrepresentation in special education the issue has been addressed through the acknowledgment of systematic bias in selection procedures, which results in disproportionality. Barona and Faykus (1992) found that discrimination in eligibility placement occurred based on various sociocultural variables, specifically, ethnicity and socioeconomic status. The authors conducted a study to examine the specific role of a set of sociocultural variables (socioeconomic status, ethnicity, familial size, father absence) on special education eligibility categories. Three hundred students from a large metropolitan school district in the Southwest who were previously placed in three categories (mental retardation, learning disability and not eligible) were randomly selected as participants. These students represented African American, Caucasian and Mexican American ethnic heritages.

Analyses of variance and multiple regression analyses were conducted to determine whether the three ethnic groups differed categorically based on the various sociocultural factors. The results in the three categories indicated that Caucasian families had a significantly smaller family size than both the Mexican American and African American participants. Additionally, the mean percentage of father absence was 64 for the African American group which was significantly greater than either the Mexican American or Caucasian groups. Finally, the Caucasian participants had significantly higher socioeconomic status scores than the two minority group samples.

Using the multiple regression analyses, sociocultural factors accounted for four percent of the variance in placement decisions for the mental retardation group. However, independently, socioeconomic status and ethnicity made significant contributions to the prediction of mental retardation eligibility. Socioeconomic status alone accounted for 4 percent of the variance and ethnicity accounted for 1.23 percent of the variance for the mental retardation category. Sociocultural variables combined accounted for total of 5 percent of the variance in the learning disabilities group. Both ethnicity and socioeconomic status made significant contributions to predictions of this group, meaning ethnicity accounted for 3.48 percent of the variance and socioeconomic status account for 1.10 percent of variance. In the group identified as not eligible, sociocultural variables combined contributed to 5 percent of total variance, a statistically significant contribution. In summary, variance accounted for by all sociocultural factors including ethnicity were small but significant with socioeconomic status and ethnicity retaining most of variance. These findings thereby provide some support for an attribution of sociocultural variables to the process of eligibility decision making.

School personnel bias in referrals. It has been proposed that the bias can occur in a test (Gould, 1981; Medina & Neill, 1990), or in the testing procedures (Fuchs & Fuchs, 1989a) or in the classroom teachers who recommend children for testing are biased (Algozzine, Christenson, & Ysseldyke, 1982; Algozzine, Mercer & Countermine, 1977). Bahr, Fuchs, Stecker & Fuchs (1991) investigated the latter because it was believed that of these three biases, the most pivotal in referrals of individuals to special education placement is possibly teacher bias because the majority of teacher referrals often lead to testing which in turn lead to special education placement (Algozzine, Ysseldyke, & Christenson, 1983).

Bahr et al. (1991) studied what types of students were more likely to be referred for an evaluation towards placement in special education. In this study, 48 general education teachers (33 white, 15 black) from 9 inner city middle schools in a southeastern metropolitan city participated in nominating 40 students (20 black, 20 white) who were not identified with a disability, but were judged by the teachers as the most difficult to teach and at risk for referral in special education placement. The dependent variables included background information (retention), teacher descriptions of target behaviors, rating scales of student behavior towards academic progress, and reading achievement as measured by two subtests of the Woodcock Reading Mastery Tests.

The results showed a statistically significant effect for background information [$\chi^2(3, N=38) = 8.73, p<.05$] meaning that black students had been retained more frequently than white students. In fact, the authors found that 13 out of 18 black students had been retained at least once, while only 5 out of the 20 white students had experienced

grade retention. There were no statistically significant differences in teacher descriptions of the target behavior. However, statistical significance was found in teacher ratings of student behavior toward academic progress. Both the black and white teachers reported more negative comments regarding the black students than towards the white students [F (1,28) = 5.29, $p<.05$]. On the Woodcock Reading Mastery subtests, the white students scored statistically significantly higher on the Passage Comprehension and Word Identification tests [F (2,37) = 3.48, $p<.05$].

Therefore, a central finding of this study was that a significantly larger percentage of "at risk" black students were rated more appropriate for referral by black and white general education teachers. The authors noted that both black and white teachers appeared to perceive black and white students classroom behavior as similar which did not appear to be a representation of the basis for teachers tendencies to perceive black children as more appropriate for referral.

However, the more likely basis for the differential perception was probably based on the academic performance of the two groups because black students scored lower on the Woodcock test and more black students had been retained at least one grade. The authors concluded that it appeared that the teachers in this study viewed black children as more appropriate for referrals on the grounds that they tended to be weaker students in greater need of specialized instruction. The authors did acknowledge that there were limitations to the study. First, the small sample sizes may have construed the results of the study as well as the fact that all the teacher participants were volunteers in a related study. Finally, the students reading achievement were estimated using two subtests of a published norm-referenced standardized assessment measure, therefore, the nature of the

academic achievement data was limited. Additionally, the participants were presented with just the Woodcock Reading Mastery scores and not more intellectual or achievement data. While the authors conducted a meaningful study, the conclusions drawn lead to more questions than explanations. The authors were unable to address are the reasons why the African American participants were consistently retained in a least one grade and had obtained lower test scores.

Research conducted by Ysseldyke, Algozzine, Regan, & McGue (1981) may provide more information regarding what influence test scores and naturally occurring student characteristics have on the psycho-educational decision making process. In this research study, 159 special educators, teachers, and school psychologists were asked to read a case folder description of a child. After reading the description, the participant could view a diagnostic simulation computer program that contained the following information: IQ scores, academic achievement scores, perceptual motor scores, personality scores, language test scores, adaptive behavior scale scores, behavioral observation results and behavioral checklists. All of the scores on the computer program indicated average results. The participants were randomly assigned to 16 conditions based on the physical characteristics of the child (sex, socioeconomic status, physical appearance, and type of problem). The researchers were interested in answering the following questions: which kinds of assessment data were used as a function of referral information; to what extent did specific student characteristics bias outcome decisions; to what extent did the participants perceive different kinds of assessment data as influencing their decisions; and to what extent did the participants perceive that naturally-occurring events influence their decisions.

The results showed that intelligence and academic achievement test scores were selected most frequently by the decision makers. The authors reported significant main effects for the referral question of emotional disturbance. Meaning, if the reason for referral was emotional disturbance, the student was more likely to be referred for a special education placement. The participants perceived that intelligence scores, academic achievement scores and the ability-achievement discrepancy had the most impact on their decisions and that socioeconomic status had the least amount of influence on their decisions when the student was from a low socioeconomic background. In summary, the authors found the following regarding the behaviors of decision makers: 1) they do look at socioeconomic status, physical attractiveness and the reason for referral; 2) intellectual and academic achievement scores were used in the decision making process; 3) the reason for referral had the most influence on final outcome; 4) when the students' socioeconomic status was high, it was very influential; and 5) students referred for behavior problems were diagnosed most frequently with emotional disturbance. The central findings of this study were two fold. First, there is a tremendous amount of weight placed on the reason for referral and school psychologists must be cautious to not enter a "confirmatory assessment" process. Second, there is a heavy reliance on test scores in the determination of special education placement.

While the conclusions of the research do suggest that bias may exist in the referral process, there are limitations to the study. First, the results of the statistical analyses were not presented in the article, which makes it difficult to substantiate the results. Secondly, the data given to the participants were computer simulated which may not reflect true

student outcomes. Finally, the participants were all located in the state of Minnesota, which may limit generalization.

From these two studies it is apparent that the special education decision making process is influenced by the reason for referral and test scores available to the decision maker. However, it has been suggested that additional "within person" variables impact the decision to refer a student for special education placement. Soodak & Podell (1993) investigated the relationship between teachers' self perceptions of effectiveness and their preconceived beliefs regarding students based on socioeconomic status and the etiology of their learning problem have an impact on the decision to refer for special education evaluation. 240 regular education teachers who had taught in the New York public school system for a minimum of one year were participants in the study. Participants were asked to read a case study regarding a male student with good behaviors, reading and concentration difficulties. All cases were identical except that the etiology for the learning problem was identified as medical, environmental or unspecified. Additionally, the student also was identified as having either a high or low socioeconomic status.

The results showed a statistically significant correlation ($r=.31$, $p<.002$) for teachers' placements and referral judgements. Specifically, there were statistically significant results for the impact of personal efficacy ($F=8.80$, $p<.01$). A statistically significant interaction was found between personal efficacy and socioeconomic status. Statistically significant main effects also were found for teacher efficacy ($F=6.71$, $p<.01$). These results indicated that teachers with a high personal efficacy tended to think that the regular education setting was appropriate for lower socioeconomic status students. The opposite also was found to be true. Additionally, there was a greater tendency towards

referral when there was an unspecified etiology. However, there were limitations to this study. First, the researchers lacked control in the procedure. The case studies were given to the participants and asked to be returned at their discretion. The sample selected consisted of all students enrolled in graduate training programs at a New York university. Again, the case study was simulated data and may not generalize to true student outcomes.

Nevertheless, the central findings of this study support the data gathered thus far in that special education referral decisions may be based on variables other than the student educational needs. Teacher personal and professional efficacy influenced their decision in referring students for a special education evaluation. The research up to this point suggests that students may be treated inequitable in the special education process based on many different variables.

As previous mentioned a moderator variable to the special education referral process may be the physical characteristics of the student. Andrews, Wisniewski, & Mulick (1997) investigated the influence of age, height and weight on teacher referral decisions. Using two samples of children previously identified for a psycho-educational evaluation, the authors examined the influence of birth month, age, height and weight in teacher's decisions to refer students for developmentally delayed and behavior disordered special education evaluations. The sample contained 213 students (47% Caucasian, 62% African American, 2% other and 2% missing) who were referred for special education evaluations. The results indicated a statistically significant effect for race in developmental handicapped referrals (χ^2=8.29, p<.02). Meaning, African Americans were overrepresented in referrals for developmentally handicapped. There were no

statistically significant effect found between IQ score and birth month. However, boys were more likely to be referred for a behavior disorder than girls (χ^2=59.6, p<.001). Statistically significant results were found for height and weight in both special education categories. Meaning that the referral rate was higher when the teacher presumed that a student's height and weight were above average for their age. The authors used archival data in conducting this research which may be a threat to validity of the results. Additionally, experimenter bias may have influenced the selection of the student to be included in the research. This study lends support the previous research discussed which stated that referrals for special education evaluations are moderated by many variables. Teachers and school personnel may be inaccurate referral agents.

In summary, it appears that students experience bias in the early stages of the eligibility process. With the overrepresentation data presented, it can be concluded that African American students are referred more frequently for special education evaluations. While the classroom teachers are charged with the initial responsibility to refer students for possible placement, federal law requires each referred student to receive a psycho-educational evaluation by a school psychologist. Therefore it is necessary to investigate the role that school psychologists play in the making the special education recommendation.

The role of school psychologists. School psychologists play a significant role in assisting in determining eligibility for special education placement. The ideal scenario is for school psychologists to follow an objective hypothesis testing procedure to determine the eligibility of a student for special education services. Instead researchers have found that hypothesis testing is subject to "behavioral confirmation" (Snyder, 1982), "belief

perseverance" (Ross, Lepper & Hubbard, 1975), and "schema driven information-processing" (Taylor & Crocker, 1981). These supplementary influences suggest that the initial impressions, preliminary considerations and hypotheses testing process, systematically bias the assessment process (O'Reilly, Northcraft & Sabers, 1989).

O'Reilly, Northcraft, and Sabers (1989) conducted a study to examine the effect of the special education referral on the interpretation of assessment data aspect of the school psychologists' role in special education eligibility decisions. Each school psychologist was provided a typical psycho-educational report of assessment data collected when evaluating a special education candidate. The authors' first research hypothesis was that the eligibility decision would be consistent with the reason for referral. The second research hypothesis was that the school psychologists would differentially weigh the importance of assessment information and recall assessment information in a referral consistent manner.

Forty-three practicing certified school psychologists employed in seven Arizona school districts volunteered to participate. Each school psychologist evaluated a simulated psychological report. All reports were identical except for the reason for referral, which served as the independent variable for the study. The reason for referral reflected typical qualities of children with either a learning disability or gifted tendencies. Each psychologist was given time to review the report, return it to the experimenter and then answer five short questionnaires regarding the report.

A Fisher's exact test indicated that classification choice was significantly influenced by referral. Furthermore, the recommendations of LD classification were larger than the proportion of gifted referral school psychologist recommending doing so.

The reason for referral represents a potentially biased influence on the special education placement process. There was a significant tendency for eligibility judgments to mirror the stated reason for referral. Additionally, the school psychologists acknowledged the importance of the reason for referral in the decision making process, however systematically discounted it as a primary agent in determining eligibility. One weakness of this study was that the summary statistics for the Fisher's Exact Test were not provided. However, the authors did provide results of follow-up t tests. For the learning disability referral condition, the authors reported statistically significant scores, $t(28)=3.25$, $p<.004$, but not for the gifted referral condition. Additionally, the small sample size (N=43) may further limit the generalizability of the results.

In summary, further studies must be conducted to assess variables that may impact the hypothesis testing process other than the actual scores and observations obtained during the testing procedures. While there were many variables such as the school psychologist, teacher opinion and test bias that may influence the rate of referral of African American students to the special education process, the African American community does recognize these differences and has chosen to respond with legal representation in the court system to rectify these ills.

Special Education Placement and Evaluation Assessments

The literature reviewed thus far has found that there may be possible bias in the special education decision making process based on many variables. Frequently, referral agents review intelligence and academic achievement scores during the decision making process. However, is it a "safe" assumption that the scores obtained from these instruments are valid to be included in the process? As it previously has been stated,

intelligence and academic achievement scores are a federally mandated part of the evaluation process. Therefore, it is necessary to investigate their utility in the special education placement process.

It has been asserted by many that intelligence and academic achievement tests provide valuable information to the special education placement process. Kaufman (1994) claims that, "children who are referred for evaluation have problems, and we can help solve those problems by interpreting IQ tests intelligently" (p.26). However, does research exist to defend the supposition that intelligence tests are useful in the process? Gresham and Witt (1997) argue that, "intelligence tests contribute little reliable information for the planning, implementation, and evaluation of instructional interventions for children and youth" (p. 250). The authors assert that the major reason why intelligence tests do not contribute to the process is because there is not a sufficient body of literature to support the existence of an aptitude by treatment interaction. An aptitude by treatment interaction (ATI) is the belief that the measurement of aptitudes (individual characteristics or traits) can predict the probability of success when given certain treatments (specific education programs). In their review of the literature, the authors found three models of ATI used to attempt to identify the relationship between underlying cognitive process and academic or instructional methods. First, the modality matching model attempted to identify inherent strengths and match instructional methods to enhance these strengths. Second, the cognitive style or processing model uses the same logic as the modality matching model accept that the strengths are identified in terms of cognitive process (e.g., simultaneous/successive, reflective/impulsive). Finally, the neuropsychological model uses brain functioning as the underlying aptitude and finds

parallel strengths (e.g., left/right hemisphere problems & strengths). While each of these models seem to provide a basis for using IQ tests or other assessment measures to identify and "solve" problems, the research does not support such conclusions. Research has failed to show consistent results of modality matching (Arter & Jenkins, 1979; Kavale, 1990; Kavale & Fortness, 1987; Ysseldyke & Mirkin, 1982); cognitive style/processes (Ayers & Cooley, 1986; Ayers, Cooley, & Severson, 1988; Das, 1995; Das, Naglieri, & Kirby, 1995; Good, Vollmer, Creek, Katz, & Chowder, 1993); and neuropsychological process in identifying appropriate aptitudes by treatment (D'Amato, 1990; Hartlage & Terzrow, 1986; Reschly & Gresham, 1989). Therefore, the assumption that there are plausible treatments that can be inferred from scores on assessment instruments is yet without support in the literature.

Additionally, there also is a dearth in the literature regarding the relationship between academic performance and scores obtained from intelligence tests. IQ has been shown to be a weak predictor in longitudinal studies of reading acquisition (Siegel, 1989, 1992; Share, McGee, & Silva, 1989, 1991). While many studies exist regarding the reliability of various intelligence tests with academic achievement measures, the question still remains as to whether either of these assessment tools provide an accurate picture of student performance capabilities.

In fact, Hall (1985) investigated the technical adequacy of 37 commonly used achievement tests. His results indicated that the majority of tests (81%) published normative data more frequently than any other type of technical data. Less than half (41%) reported some type of criterion related validation data and slightly over half (54%) provided information regarding where how test items were obtained. 73% of the tests did

report content validation data, however, not all of the tests identified the procedures by which the skill domains were derived. The most interesting information was that only 33% of the tests reported using a set of procedures to minimize ethnic and gender biases. Additionally, only 8% of the instruments reported using Rausch item response scaling as a method to select the most unbiased items. From the Hall study it is apparent that academic achievement measures also may be lacking in technical adequacy. Over the past twenty years additional research has been conducted that shows a widening gap between what skills are taught in the public school system and the content tested on standardized achievement tests (Jenkins & Pany, 1978; Shapiro & Derr, 1987).

Differential outcomes in performance for African American students also have been found in academic achievement measures. Glutting, Kelly, Boehm, & Burnett (1989) examined the performance of African American kindergartens on the Boehm Tests of Basic Concepts- Revised (BTBC-R). The BTBC-R was described by the authors as a measure of basic skills to assess kindergarten readiness. The sample consisted of 58 children all enrolled in a kindergarten program (1985-1987) in the northeast region of the United States. The children were given two criterion related achievement measures, the Metropolitan Readiness Tests-level 2 and the Cognitive Skills Assessment Battery-second edition. The students were given the BTBC-R once in October and re-assessed in April. The results obtained showed that the African American children scored "somewhat below average." Also, there were statistically significant correlations found between the BTBC-R and the CSAB (r=.82) and the MRT (r=.60). The test-retest correlation for the BTBC-R also was statistically significant (r=.82). The authors concluded that because the test obtained high intercorrelations with other frequently used achievement measures and

because of the high test-retest reliability, the BTBC-R was an adequate measure for use with the African American kindergarten population.

A limitation to the generalization of the author's conclusions is that the sample selection methods and size may not be an accurate reflection of the total African American population and therefore may not accurately reflect total group performance on the instrument. Additionally, many of the statistical findings were not presented therefore, limiting the ability to validate the obtained results. Nevertheless, the authors did provide evidence that there are still differences found between ethnic minority groups when comparing academic achievement scores. The authors conclusions also raise a question to the assessment community's process to insure test validity and reliability for ethnic minority populations. The supposition that if the test produces similar results as frequently used measures, then it is an appropriate assessment tool should be examined for true accuracy across ethnic groups.

While there may be differences between the standardized achievement test content and what is covered in school curricula, there is research evidence that suggests that improvements in moderator variables may not impact the differential achievement patterns found between African American and Caucasian students. Link & Mulligan (1986) investigated the outcome of providing additional time in the basic academic skills areas of reading and mathematics during the school day. From a sample of over 110,000 elementary school students, pre and post test data were gathered during the 1976-1977 school year. A comparison of the pre and post test scores showed that African American (X=28.25) and Hispanic (X=29.12) students obtained statistically significant lower scores

than Caucasian (X=39.62) students in the reading and mathematics academic achievement measures.

Additionally, the authors provided more instructional time to ethnic minority students than to the Caucasian students. Using a regression model, the authors did not find statistically significant gains in either reading or mathematics performance based on elongating the instructional time. The authors also examined the relationship between racial composition of the classroom and academic achievement (Link & Mulligan, 1991). Using the same data set, regression analyses were conducted on various models of classroom arrangements. Again, the authors found no statistically significant effects for racial composition of classrooms on academic achievement tests. Both of these studies were limited by the use of archival data. Additionally, because the student information accessed was completed in the 1970's, the results obtained may not be representative of the students of today. Nevertheless, these studies provided information regarding the impact of moderator variables on changes in achievement scores.

Even though there are problems with utility and technical adequacy of intellectual and academic achievement measures, again, they are a required element in the special education evaluation process. Therefore, it is necessary to continue the investigation towards identifying the least problematic instruments available. The majority of complaints regarding unfair test practices focus on intelligence measures. An examination of the construct validity of various instruments may provide additional assistance in identifying the most valid measures across ethnic groups.

Construct validity and the wechsler tests. The use of intelligence tests in the special education evaluation process has been scrutinized for various reasons. However,

the most prevalent complaint over the past two decades has been that intelligence tests are biased towards ethnic minority groups (Elliott & Boeve, 1987; Jensen, 1976; Reynolds, 1980; Slate & Jones, 1995; Taylor, Ziegler & Partenio, 1984; Valencia, Rankin, & Oakland, 1997; Vance & Sabatino, 1991). Some authors believe that sufficient evidence is available to suggest that intelligence tests are harmful and biased towards many African American students (Gould, 1981; Hilliard, 1991; Jones, 1988; Patton, 1992, 1998). Reynolds (1982b) stated that the issue of bias is one of the most critical issues facing current-day psychology and that this is a question that must be answered on a test by test basis and not sweeping generalizations across the psychometric field. Therefore, it is the purpose of this section to investigate the literature regard the issue of test bias in the Wechsler series.

Test bias may take many forms. Reynolds (1991) defined bias as "constant or systematic error in the estimation of test scores of an individual examinee that is due to a nominal variable such as gender, ethnicity or socioeconcomic status" (p. 21). The most frequently used methods to determine test bias begin with an investigation into the validity of a specific instrument. As previously stated, construct validity is believed to be the "overarching" form of validity (Messick, 1989). Construct validity is defined as "the extent to which a test is capable of measuring a hypothetical trait or construct" (Witt, Elliott, Kramer, & Gresham, 1994, p. 109). A test that has construct validity can demonstrate that it measures what it is supposed to measure across various ethnic minority groups (Anastasi, 1988). If a psychological test measures different hypothetical traits (e.g., intelligence) for one group than it does for another group, then the test may have construct bias (Keith & Reynolds, 1990). Therefore, in order to begin an

investigation of test bias, the literature review will focus on the presence or absence of construct validity in the Wechsler series.

In 1979, Vance, Hankins, & McGee conducted a study to investigate subtest differences in the WISC-R for a group of African American and Caucasian students. Participants were 120 students from the Appalachian area who were matched on their obtained Full Scale IQ scores and "mental age". The study included 30 students in each group that were separated by gender and ethnicity. The Full Scale IQ scores ranged from 50 to 81. The researchers conducted t test for matched pairs statistics for ethnicity and t test for independent sample statistics for gender.

The results showed that African American males obtained significantly higher Verbal IQ scores than did the Caucasian males. There were no statistically significant differences found between females. The authors conclude that the WISC-R is sensitive to establishing individual subtest differences between subjects from different ethnic populations. However, there are limitations to the study. The authors did not report the statistics obtained from the t tests. Additionally, the authors stated that they did not provide corrections for the error generated from computing over 30 t test comparisons. Finally, the use of t tests did not employ the most sophisticated statistical analysis tools available to determine if differences actually existed.

Also in the 1970's, Munford (1978) investigated the performance of African American psychiatric patients with the WISC-R. Ten African American males and ten African American females were given WISC-R tests as a part of their psychiatric re-evaluations. The author found that African American males obtained significantly higher scores on the Verbal ($F=32.57$, $p<.0001$) and Full Scale ($F=21.87$, $p<.001$) IQ scores

when compared to African American females. These results must be interpreted with caution because of the small sample used in the study.

In a factor analytic study of the WISC-R conducted by Taylor, Ziegler, and Partenio (1985) a two-factor model (Verbal & Performance scales) yielded identical results for African American, Caucasian and Hispanic students. The three-factor model (Freedom from Distractability, Verbal Comprehension & Perceptual Organization) yielded similar loadings for the Hispanic and Caucasian samples, but not for the African American students.

Finally, Reynolds (1980) used the standardization sample data from the WISC-R to investigate differences between African American and Caucasian students. There were 938 Caucasian males, 137 African American males, 927 Caucasian females and 153 African American females represented in the samples. The author then contrasted the highest and lowest correlations between the four groups to determine if the relationship between age and performance was equivalent across groups. Then regression coefficients were calculated between age and raw scores across groups. The results of the correlation comparisons indicated that there were no statistically significant differences between the two groups. However, the regression coefficients yielded smaller incremental changes for African American males than for the other groups. Meaning, African American males did not make the expected Full Scale IQ score gains that were predicted of that group. The author contends that, "the rate of development with respect to "g" is not equivalent across race and that African Americans (males in particular) develop these skills at slower rates that their Caucasian counterparts" (p. 377). The limitation in generalizing this conclusion is that the author did not provide the statistical support for such a claim. Additionally, the

regression coefficient difference found was the difference found between the groups in over 30 comparisons made.

While there were early studies that demonstrated differences in ethnic groups on the WISC-R, numerous studies also found no differences in ethnic group score patterns. Vance & Engin (1978) used hierarchical factor analytic methods to investigate the intercorrelation matrix of the WISC-R for 150 African American students who had been referred for special education evaluations. The authors found that the two-factor model yielded highly similar results as reported in the standardization sample from the WISC-R. Additionally, Gutkin and Reynolds (1981) found that the factor structure of the WISC-R was essentially identical when comparing a group of African American and Caucasian students drawn from the WISC-R standardization sample data. Sandoval (1979) studied the internal consistency of the WISC-R subtests for a group of Mexican American, African American and Caucasian students and found the internal consistency estimates to be within .04 of one another. Juliano, Harrad, & Carroll (1988) studies the factor structure of the WISC-R for over 322 students classified as learning disabled. The results indicated that the WISC-R factor structure was stable over the three-year time span and there were no differences based on gender or ethnicity. In summary, there has been conflicting evidence over the past twenty years regarding the construct validity of the WISC-R test for the African American population. In 1991, the Psychological Corporation released an updated version of the Wechsler test. The Wechsler Intelligence Test for Children-Third Edition (WISC-III) quickly became the most popular instrument in the special education evaluation process (Reschly, 1997). However, because the differences between the WISC-R and the WISC-III were foundational and not cosmetic

(Bracken, 1993), research conducted with the WISC-R cannot automatically be assumed to be generalizable to the WISC-III (Slate & Jones, 1995).

In the WISC-III manual, separate technical data information for minority group members were not report. Therefore the extent to which the reported factor structures generalize to ethnic minority groups is unknown and must be investigated. Slate and Jones (1995) conducted a study to investigate the factor structure of the WISC-III for African American students undergoing special education evaluations. The sample consisted of 58 African American students from the Mississippi delta region of northeast Arkansas who had been referred because of academic difficulties. Confirmatory factor analyses were conducted to investigate the stability of the WISC-III factor structure. Initial factor loading results indicated that 44% of the variance could be accounted by one variable. All subtests loaded on this factor at .4 or higher, thus indicating support for the Full Scale IQ. Similar support was found when varimax rotations were conducted for the Verbal and Performance IQ scores. However, there were problems during the rotation. Two subtests, Arithmetic and Digit Span did not load with the other Verbal subtests following rotation. Additionally, the Coding subtest also did not load on the Performance scale after rotation. Thereby providing evidence that there is a difference between language and numerical abilities with the African American population. Nevertheless, the generalization of these results is limited because of the small sample sizes used. It is suggested that factor analysis methods be used with samples of 100 or greater. The interpretation of these results is also cautioned in that the sample came from a restricted geographical region.

In 1997, Kush and Watkins conducted an exploratory factor analysis of the WISC III for a population of 161 African American students who had been identified to receive special education services. In order to simulate the analysis techniques that were reported in the test manual, a varimax rotation was conducted for all factors with eigenvalues greater than 1.00. The results reported were that a substantial proportion of the variance was accounted for by a general factor "g" (55%). Additionally, the authors found that the traditional factor structure of the WISC III was substantiated by the results obtained. All the subtests loaded highly on two rotated factors. In the generalization of results it must be noted that the concern for construct validity can not be determined by this one analysis, yet the results provide some evidence of the validity of this measure. Additional limitations were that participants' Full Scale IQ Scores were 20 points lower than the average IQ scores reported in the test manual; archival test records were the data pool; and all of participants had undergone special education evaluations. While these occurrences in the data limit the generalization of the results, the information gathered provided a contrast to the Jones & Slate (1995) study as to whether the WISC III has sufficient construct validity for African American children.

Multi-sample confirmatory factor analysis and construct validity. From the studies previously discussed, it is evident that researchers frequently use the factor analytic method to evaluate the validity of instruments. Factor analysis has been one of the most commonly used statistical methods to investigate the construct validity of assessment tools for different ethnic minority populations. In fact, Reise, Widaman, & Pugh (1993) examined whether confirmatory factor analysis or item response theory would be an appropriate statistical method to examine individual test scores drawn from

two different populations. The authors compared mood ratings of undergraduate students from the University of Minnesota in the United States with undergraduate students from the University of Najing in China. The independent variables were the geographic location of the students, ethnicity and culture. The dependent variable was their obtained score on the mood scale.

Both confirmatory factor analysis and item response theory procedures were used to analyze the data obtained from the covariance matrices of the groups. A model of fit was estimate for the assessment results. In the confirmatory factor analysis, the researchers used a three-step process to test for differences in performance of the groups on the scale. In the first step, the latent variables were constrained to assess full and partial measurement invariance in factor loadings. In the second step, the factor variances were constrained to be equal. In the final step, one sample's variances were constrained to allow for comparisons in the second sample group. In the item response theory, the analysis was conducted by a recalibration of theta estimates to investigate practical fit of the data.

The results obtained showed that the baseline model provided an adequate explanation of the observed data (χ^2=74.84, p<.001). However once the researchers constrained all elements to suggest full model fit, there was a statistically significant decrease in fit ($\Delta\chi$2=15.19, p<.001), therefore the researchers were able to reject the null hypothesis that there was invariance across the two groups. The results obtained from the item response theory analysis also revealed similar trends in the data. The authors concluded that the processes of using confirmatory factor analysis and item response theory were relevant to investigate of construct validity to determine measurement

invariance across groups. The confirmatory factor analysis procedures were determined to be more user friendly than the item response theory methods, however caution would be needed to rely on both the practical and statistical indices to determine statistical significance. The authors did caution that there are some who believe that the use of modification indices leads to post hoc "fitting the data." Additional research has been conducting that provides support to the use of confirmatory factor analysis to investigate measurement invariance across groups.

Byrne, Shavelson, & Muthen (1989) demonstrated the application of confirmatory factor analysis in testing for group mean differences. Using data gathered from 11^{th} and 12^{th} grade low and high ability tracked participants in the development of a self concept measure, the researchers searched for differences in the low and high ability tracked students using confirmatory factor analysis. First, the authors examined the data to establish baseline models. Next, the invariance of the self-concept measurements and structure across the groups was investigated. Finally, the latent mean track differences were tested for equality while providing constraints on the factor variables.

The results showed that the hypothesized model of fit was not an adequate representation of the obtained data (χ^2=160.54, BBI=.89, TLI=.87 low track and χ^2=401.09, BBI=.92, TLI=.89 high track). Therefore the hypothesis of equivalence of covariance matrices also was rejected. While the factor loadings were not found invariant, there were differences found in various subscale areas like general academic, English and mathematics self-concept. In conclusion, the differences found in the overall general self-concept were not invariant across the high and low track students. Limitations of this study include that there was not an in depth description of the samples

provided and the sequential testing of models in the exploration of partial measurement invariance may have influenced the results obtained. Nevertheless, the study does provide support for the use of confirmatory factor analysis in investigating measurement invariance across groups.

Within the school psychology literature, the use of multi-sample confirmatory factor analysis (MCFA) has been demonstrated in the examination of measurement invariance for the WISC III (Keith & Witta, 1997; Kush, Watkins, Ward, Ward, Canivez & Worrel, 1999). Keith & Witta (1997) used MCFA techniques to answer the questions of whether the WISC III measures the same constructs over the 11-year age span of the test. Through an analysis of the WISC III standardization data correlation matrices and standard deviations as reported in the test manual, the authors engaged in a stepwise confirmatory factor analysis process to examine the covariance matrices, factor structure, factor loadings and variances of the test. From the results obtained, the authors failed to reject the hypothesis that the WISC III measures the same construct across age groups ($\chi^2(671)=553.06$, $p>.05$). There was also evidence that the factor structures were equivalent across age groups ($\chi^2(971)=732.67$, $p>.05$). Next, the authors found that the factor loadings and subtest unique and error variances were equivalent across the 11-year age span. However, the authors did find that a four-factor model was a better fit to the data than the two factor model suggested in the test manual. While the authors concluded that the WISC III appeared to be a consistent measurement instrument across time, they warned the practitioner in interpreting a two factor instead of a four-factor model. Even though this study was based on the covariance matrices listed in the test manual; the

authors did provide evidence regarding the use of multi-sample confirmatory factor analysis in determining the construct validity of an instrument.

While the literature discussed thus far regarding the use of confirmatory factor analysis has focused on different countries, placement tracks and age groups, Keith, Fugate, DeGraff, Diamon, Shadrach & Stevens (1995) provided the seminal work regarding the use of multi-sample confirmatory factor analysis in investigated construct bias based on ethnic differences within the United States. This research study demonstrated the use of multi-sample confirmatory factor analysis to study construct bias in the Kaufman Assessment Battery for Children (K ABC). The sample consisted of raw data obtained from 813 Caucasian and 486 African American students tested during the K ABC standardization process. The first step in the multi-sample confirmatory factor analysis procedure was to develop correlation matrices for each age level and ethnic group. The hypothesis was to specify these matrices to be equal for both groups. Second, the factor structure as explained in the test manual was tested for equivalence between the groups. This second procedure included an investigation of the factor loadings, the factor correlation, factor variances and subtest unique and error variances were tested for equivalence across groups.

From the results obtained, the K ABC was determined to be a relatively bias free instrument. The authors did not find statistically significant differences when examining the differences between the covariance matrices of both groups. However, statistically significant differences were found for the 9-10 year old groups when the subtest unique and error variances were not allowed to vary ($\Delta\chi^2$=25.69(13), p=.019). Statistically significant differences also were found for the 11-12 year old group when factor loadings

and correlations were invariant ($\Delta\chi^2$=10.73(3), p=.013). While statistically significant differences were found in these areas, the differences were due to less critical aspects of measurement. The authors concluded that the K ABC appeared to measure the same thing across the age groups, however, whatever is being measured may not be what the test publishers suggest. In brief, the school psychology literature has begun to examine the use of measurement invariance investigation procedures to determine the construct bias of frequently used assessments. The focus of the current study is to extend the literature base by examining construct bias in the most frequently used IQ test, the WISC III, for ethnic minorities students.

Summary

In summary, the literature review began with a discussion of the history and trends of education and African American students. It was discussed that African American students have been thought to achieve intellectual and academically below other ethnic groups. The misconception of underachievement led to and was reinforced by systematic discriminatory practices such as ability grouping, tracking and overrepresentation in educable mentally handicapped special education programs. The response of the African American community to such practices was litigation. The review then examined court cases that have sought to improve the educational plight of African Americans. With the most controversial issue being the overrepresentation of African American students in educable mentally handicapped classes, the special education process was discussed next. The roles that teachers, school personnel and school psychologists play, from the referral through the assessment given, were crucial to the inquiry of why African Americans experience differential educational outcomes in the

public school environment. To further investigate the trend of overrepresentation, the review focused on the intellectual measures given and the presence of construct bias. Specifically, the WISC III was discussed because of it being the most frequently used IQ measure. Researchers in the school psychology field have begun to use more sophisticated analysis techniques to examine measurement invariance. One emergent technique has been multi-sample confirmatory factor analysis. The purpose of this research study is to conduct a multi-sample confirmatory factor analysis of the WISC III to determine measurement invariance between African American and Caucasian students.

Chapter 3

Method

The purpose of this study was to compare the factor structure of the WISC III for a sample of African American and Caucasian students. This chapter begins with a discussion of the students selected to participate. The questions of who would participate, how many would participate, and how they would be selected were answered. The next section outlined the measures used. Specifically, information regarding the validity and reliability of the WISC III are provided. The procedures used to collect the data and the data analysis process involved in multi-sample confirmatory factor analysis are outlined.

Participants

The participants were selected from the population of students age 6 through 16 who were identified for assessment (academically or behaviorally) in the Hillsborough County (Tampa, Florida) public school system from the 1994-95 through the 1998-99 school years. The Hillsborough County school system, the 12th largest school district in the United States, has an enrollment of approximately 154,000 students (Hillsborough County Government, 1999). The reported ethnic stratification of the district is as follows: Caucasian 75.4%, African American 12.2%, and Hispanic 12.4%. Hillsborough County reported a teacher to student ratio of 1 to 16.5, which was described as "one of the lowest" of the thirteen largest school districts in the nation (Hillsborough County Government, 1999).

Referral Process

Student IQ score data were accessed from the Hillsborough County Public Schools central files database. The database consists of student data that were collected during the referral process for special education eligibility determination. The referral process for students in Hillsborough County begins by the classroom teacher completing a request for assistance to the Intervention Assistance Team because a student is demonstrating either superior or insufficient academic or behavioral achievement. The Intervention Assistance Team recommends interventions that the teacher can use with the referred student. If satisfactory improvement is made, the process ends. If satisfactory progress is not made, the student is directed to the Child Study Team, which may include the classroom teacher, school guidance counselor, principal, assistant principal, special education specialist and/or school psychologist. Any member of this team may complete an official referral for psychological testing or special education assessment. After the referral for testing is completed, the referral is sent to central files to be entered into the computer system. A copy of the referral data screen can be viewed in Appendix I. The referral data screen lists the student name and number, the reason for the referral, the referral initiation date, the referral completion date, the school number and the work number of the person submitting the referral. The referral is then sent to the Psychological and Diagnostic Services Department where it is assigned to a state-certified school psychologist. The assessment process follows:

1. The cumulative records of the referred student are reviewed.
2. The referred student is observed in the classroom setting.
3. A student interview is conducted prior to the assessment.

4. The student is given a battery of tests depending on the reason for referral. Most students are given intellectual and academic achievement measures.
5. The results of these assessments are written in a report that is reviewed by the school Child Study Team. The Child Study Team uses the data generated from the report in conjunction with other data to determine special education placement eligibility.
6. The assessment results are logged into the school district central files database. A copy of the assessment data screen is included in Appendix II. The database lists the student name and number, the central file number, date the test was given, the type of test and the score earned. The school psychologist recommendations regarding special education placement also are logged into the school district central files database (see Appendix III). The recommendations screen lists the student name and number, the central file number, the diagnosis code and the date.

The Hillsborough County Public School System reported that a statistically significant number of African American males are overrepresented (i.e., higher referral percentage than what would be expected based on their population in the school district) in the referrals for behavioral difficulties, however these data were determined to not be clinically significant (B. Haines, personal communication, September 20, 1999). The school district reported a "perceived" underrepresentation (lower referral percentage than what would be expected based on their population in the school district) of minority students in the gifted special education category. However, with the use of an alternative identification process for minority students, there are comparable numbers of students of all ethnic groups represented in the gifted program. During the 1999-2000 school year, the school district began to require the teacher and referral agent to report their own

ethnicity on the referral form. This researcher attempted to gather data for the number of African American students annually referred for special education evaluation and the number of African American students annually referred to the gifted education program, but they were not available from the Hillsborough County Public School System. However, from the number of students represented in special education programming, it is assumed that there is a difference in the referral rates of students by ethnicity. Of the total school district population, there are 30,251 (approximately 19%) students enrolled in special education programs (Florida Department of Education, personal communication, October 15, 1999). Caucasian students account for 75% of the total school district population and 57% of the special education population. African American students account for approximately 12% of the total school district population and 23% of the special education population. Hispanic students also account for approximately 12% of school district population and 15% of the special education population. From these data it appears that the African American population may be overrepresented in the special education population in comparison to their representation in the school district.

A "z" test was conducted in order to provide information regarding the statistical significance of the representation of ethnic minorities in the special education referral process. A statistically significant z was found ($z=-51.6$) when examining if the number of African American students in the referral sample was proportional to the number of African American students in the Hillsborough County Public School District. These results suggest that there may be a referral bias in the special education process. The African American student population may be overrepresented in the special education population in comparison to their representation in the school district.

Participant Selection Procedures

The selection procedures for the participants in this study were as follows:

1. Consent was obtained from the Hillsborough County Public School System Office of Assessment and Evaluation (OAE) to conduct the investigation. The OAE requested that only persons employed by the school district access the central files database of potential participants and the central files folders of the identified participants. Additionally, consent was obtained from the University of South Florida Institutional Review Board to conduct the study. Confidentiality was maintained regarding personal student data by maintaining the listing of student names, numbers and central file numbers in a secured area.

2. A stratified, random sample plan was used to ensure that the sample data consisted of the students who met the criteria for inclusion. The pool consisted of students who had not been identified by the Child Study Team to receive special education services and students from the special education population who had been identified with gifted, learning disabled, educable and trainable mentally handicapped disabilities. This data set included 23,254 African American, Caucasian, Hispanic, Asian, Indian and Mexican students from the Hillsborough County Public School District. Because the current study only examined African American and Caucasian students, the Hispanic, Asian, Indian and Mexican students were deleted from the pool. The final pool consisted of 18,540 African American and Caucasian students.

3. The potential participants were selected from the central files database based on the following parameters: the referral and testing occurred during the 1995-96 through 1998-

99 school years, they were of Caucasian or African American ethnicity and they were given the WISC III intelligence test.

4. The potential participants were sorted by a number of variables. First, the participant pool was sorted into 8 age categories (6-6.11, 7-7.11, 8-8.11, 9-9.11, 10-10.11, 11-12.11, 13-13.11, 14-16.11). To replicate the data procedures used in the WISC III manual (Wechsler, 1991), the age ranges from 11.0 to 12.11 and 14.0 through 16.11 were collapsed into two categories. Second, the participant pool was sorted by gender to ensure that the population was represented by an equal number of males and females. Third, the participant pool was sorted by IQ. Because the standard deviation of the WISC III is 15, the IQ scores were sorted into 7 categories (<55, 55-69, 70-84, 85-99, 100-114, 115-129, =/>130). Finally, the participant pool was sorted by race to insure that there was equivalency across groups. This 8x2x7x2 model created 224 categories, which were listed on the final report.

5. From the final report, the SAS (SAS Institute, 1996) statistical data analysis program was used to apply a random number to each student. Participants were selected by taking the four highest random numbers in each category. There were some categories that did not contain four students to select from. In this case, all of the students in the category were selected. This yielded 880 students as potential participants.

6. From the list of 880 potential participants, central files numbers, date of birth, gender and ethnicity were listed separately on datasheets.

7. Data collectors from the Central Files Department of the Hillsborough County School District were trained to assist the researcher in transferring the student data from the Hillsborough County central files folder into the research study datasheets (see Appendix

V). First, the central file number was obtained from the datasheets. Second, the central file was located and reviewed for WISC III data. Third, the WISC III data (e.g., Full Scale IQ and subtest scale score data) were recorded on the datasheet. Finally, the handwritten datasheets were transferred by three (3) persons to the computer database. The data transfer personnel obtained 100% accuracy in data transfer and interrater agreement prior to the start of data transfer.

8. The WISC III data obtained included the Full Scale IQ score and the scale scores (Information, Similarities, Arithmetic, Vocabulary, Comprehension, Digit Span, Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding). The Digit Span subtest is a supplementary subtest and was not consistently administered within the population. Therefore, the scores obtained from the Digit Span subtest were not included in the present study.

9. Interrater agreements were computed for 20% of the data transfers. Reliability estimates were computed by dividing the total number of accurate transfers by the number of accurate plus inaccurate transfers multiplied by 100, with a criterion of 95% accuracy. Interrater agreement equaled 100%.

Measures

The WISC III is a test that purports to measure a child's intellectual functioning based on their skills in completing various tasks requiring the use of both verbal recall of information and facts and the manipulation of objects. A composition of 10 subtests yields an overall Full Scale IQ score. The test is divided into two scales, Verbal and Performance. Tasks on the Verbal scale include answering specific academic and factual questions typically learned in school, telling how two things are alike and defining

words. Tasks on the Performance scale include the use of both speed and accuracy in physically manipulating objects. For example, the student is required to arrange pieces of a puzzle to form common objects and arranging sets of blocks to match a given pattern. The test also may be reorganized into a four-factor model of Verbal Comprehension, Perceptual Organization, Processing Speed, and Freedom from Distractibility (Wechsler, 1991).

From the information presented in the manual (Wechsler, 1991), the WISC III appears to be a reliable instrument. Split-half reliability coefficients ranged from .70 to .95 on the subtests and factor scales. 353 members of the standardization group were tested twice to assess the test-retest stability of the instrument. Given a median of 23 days, stability coefficients ranged from .56 to .94 on subtests and factor scales. Interscorer agreement was found to be adequate, ranging from .90 to .98 on various subtests. Appendix IV contains a copy of Table 5.1 from the WISC III manual. This table lists the reliability coefficients of the subtests, IQ scales and factor-based scales by age.

Exploratory factor analyses were conducted on the standardization sample data to determine the internal structure of the instrument (Wechsler, 1991). Using principal components, iterated principal axis, and maximum likelihood extractions, the test publishers found a best fit with the four-factor model. There appeared to be two major factors (Verbal Comprehension and Perceptual Organization) that account for 45% of the variance and two minor factors (Processing Speed and Freedom from Distractibility). Confirmatory factor analyses also were conducted to assess best fit to the data. While the test publishers did note the fit of the data to the younger (6-7) age groups to a five-factor model, the results showed that a four-factor model was the model of best fit across all age

groups. There have been numerous studies regarding the factor structure of the WISC III (Bracken, 1993; Carvajal, Hayes, Lackey, Rathke, Wiebe, & Weaver, 1993; Doll & Boren, 1993; Maller & Braden, 1993; Slate, Jones, Graham & Bower, 1994; Teeter & Smith, 1993; Weiss, Reifitera, & Roid, 1993; Wessel & Potter, 1994). The test manual reports WISC III Full Scale IQ score correlation with the Differential Ability Scales (DAS) General Conceptual Ability score of .92. The Verbal IQ score correlated .87 with the DAS Verbal Ability score. Finally, the WISC III Performance IQ score correlated highly with the DAS Nonverbal Reasoning Ability score (r=.78) and Spatial Ability score (r=.82). Other studies conducted found adequate convergent validity between the WISC III and other commonly used tests of intelligence, such as the Stanford Binet Intelligence Scale-IV (Bracken, 1993; Carvajal, Hayes, Lackey, Rathke, Wiebe, & Weaver, 1993).

When discussing the validity of a cognitive measure, it is important to examine how the instrument relates to academic achievement (predictive validity). The test manual reported moderate correlations between the WISC III and the Otis-Lennon School Ability Test (r=.38 to .73). Bracken (1995) reported that since the publication of the WISC III, IQ-achievement correlational studies have been conducted and found "adequate" results for normal, severely emotionally disturbed, language/speech impaired and hearing-impaired/deaf populations. For the Caucasian population, Weiss, Prifitera, & Roid (1993) found significant correlations between Full-Scale IQ scores and reading achievement scores (r=.658, p=.0001), mathematics achievement scores (r=.601, p=.0001), and writing achievement scores (r=.474, p=.0001) when the normal population of students were given the WISC III and a standardized achievement measures (e.g., Comprehensive Test of Basic Skills, California Achievement Tests). Similar results were

found with the African American population, correlations for reading achievement (r=.707, p=.0001), mathematics achievement (r=.575, p=.0001), and writing achievement (r=.499, p=.0001) were statistically significant.

When comparing adolescent males with severe emotional disturbance (ED) with a control group, Teeter and Smith (1993) found that for both groups a significant amount of the variance in reading achievement scores (e.g., WJ-R) was accounted for by the WISC III scores (ED=48%, control=72%). For mathematics achievement, the authors found 99% of the variance accounted for in the ED group and 80% of the variance in the control group. Doll & Boren (1993) examined language/speech impaired students scores on the WISC III with WJ-R scores in Reading, Mathematics and Language. While a statistically significant correlation was not found between the WISC III Full-Scale IQ scores and the obtained scores in Reading (r=.53, p>.01), there were statistically significant scores in Mathematics (r=.62, p<.01) and Language (r.=59, p<.01). With the hearing-impaired/deaf populations, Maller and Braden (1993) found adequate correlations between the WISC III Performance IQ-Verbal IQ composite and the Stanford Achievement Test-Hearing Impaired (SAT-HI) scores in Total Reading (r=.71) and Total Language (r=.77) and no statistical significance with Total Mathematics (r=.81).

Maller and Ferron (1997) investigated invariance in the WISC III when examining deaf students compared with the standardization sample. Using multi-sample confirmatory factor analysis, the authors found that the general four factor structure appeared to be similar across groups. However, the factor covariance matrices, error

variances and path coefficients were not invariant. Thereby providing data suggesting that the WISC III may measure different constructs across groups.

The test publishers also report that a bias study was conducted during the standardization process. However, the results of this analysis were not presented in the manual. In a review of the WISC III, Braden (1995) summarized that the WISC III has strong technical adequacy as reported in the manual. However, he did suggest additional areas of research in subtest stability, factor structure and the lack of a theoretical base for the hierarchical cognitive model. While the WISC III appears to be a relatively strong instrument for certain groups, the lack of information regarding testing for bias for ethnic minority groups was the purpose of the current study.

Data Analysis

The researcher used multi-sample confirmatory factor analysis (MCFA) to compare the African American and Caucasian student performance on the WISC III to examine factoral invariance. MCFA provided a more direct comparison in the investigation of factor structure equivalence across groups (Keith et al., 1995). Because intellectual measures are used with various ethnic populations, there is an assumption that the same constructs are measured across groups. Through using MFCA procedures, we were provided with more information regarding the appropriateness of such an assumption with the WISC III.

Using the EQS statistical data analysis program (Bentler, 1995), the first step began with a comparison of the two samples to examine model fit (see Figure 1).

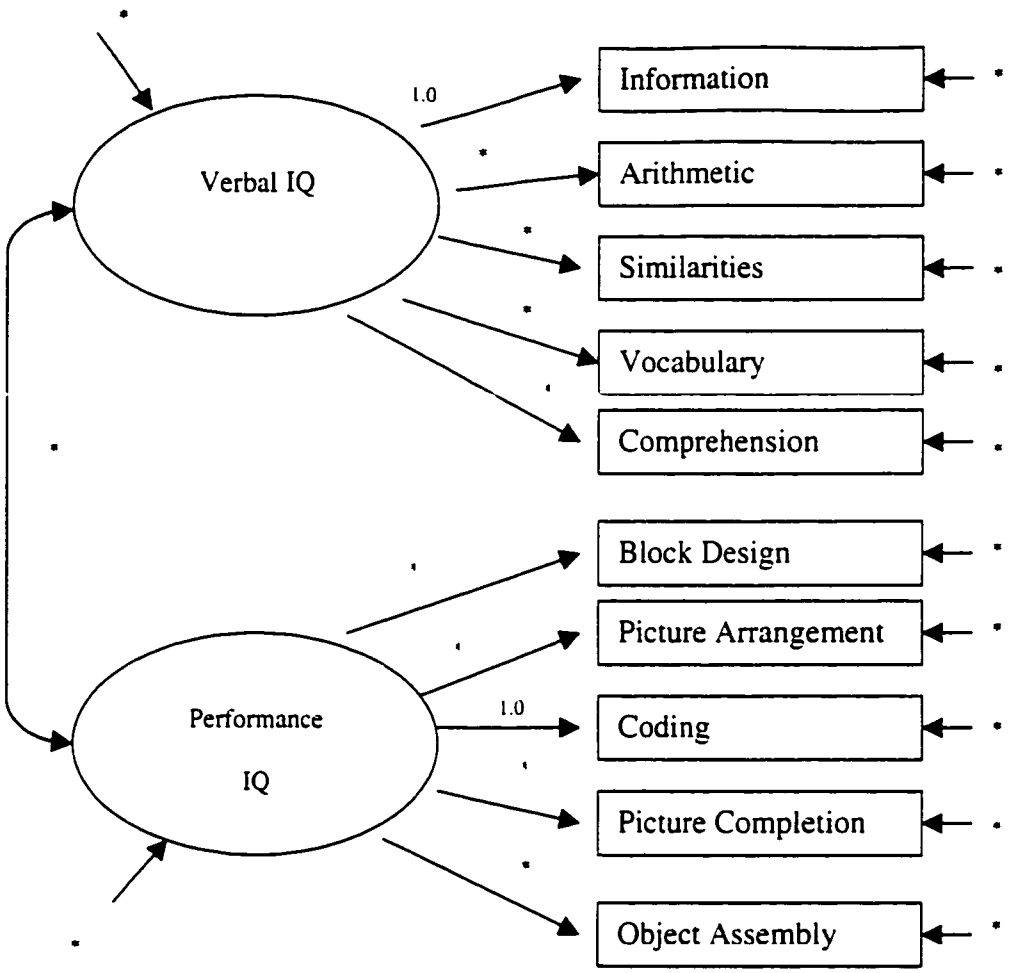


Figure 1. WISC III Two Factor Model.

All parameter estimates of the factor solution were allowed to vary. This first step was equivalent to conducting separate confirmatory factor analyses for the Caucasian and African American participants. The step procedures outline model fit in providing the most parsimonious explanation of test performance across groups. Using the analysis process suggested by Keith et al. (1995), the process continues with the researcher conducting a 4 step series of analyses during which all possible parameters (factor loadings, the factor correlation, factor variances and subtest unique and error variances) are constrained across groups. Specifically, in step 2 the factor loadings were held constant for both groups. In step 3, the factor loadings, factor variances and the factor

correlation were held constant across groups. For step 4, all parameter estimates were constrained to be equal (factor loadings, the factor correlation, factor variances, and unique and error variances). This test allows the researcher to test for fit to the specific model of the WISC III designated in the manual. The final step was to investigate equivalence of covariance matrices for the Caucasian and African American samples, without specifying a factor model. This analysis assisted in determining whether the WISC-III measures the same constructs or attributes across both groups (Witta & Keith, 1997).

The chi square was one of the measures used to determine goodness of fit. In steps 2, 3 and 4, changes in chi-square also were evaluated to determine goodness of fit. Joreskog and Sorbom (1989) describe the chi square as a measure of the fit of all models in all groups to the data from all groups. Bollen (1989) suggests that when using a hierarchy of hypotheses (such as in the current study), it is best to examine changes in chi square as parameters estimates are restricted. The changes in chi square were more appropriate because we are using nested models, meaning the hypotheses in steps 2, 3, and 4 were nested in the hypothesis in step 1. The chi square statistic is one of the most frequently reported fit statistics in confirmatory factor analysis. However, it is often supplemented by other statistics because it is directly related to sample size (Keith et al., 1995). Large samples tend to provide significant chi squares, even when a good model is specified. Therefore, to provide additional statistical support to model fit, the Comparative Fit Index (CFI) was reported. While Rigdon (1996) warns against the use of CFI in a confirmatory versus and exploratory context, the author does state that the measure may be beneficial with larger sample sizes. As an additional measure of

practical fit, the Bentler and Bonett (1980) Nonnormed Fit Index (NNFI) was used. Hu & Bentler (1998) suggest the criteria of greater than .95 to indicate appropriateness of model fit. The following research questions were answered using the step process:

1. Is construct bias present or absent in the WISC-III when comparing a sample of African American and Caucasian students? (steps 1-5)

2. Are the factor loadings invariant across the African American and Caucasian student populations? (step 2)

3. Are the factor correlation and factor variances invariant across the African American and Caucasian student populations? (step 3)

4. Are the subtest unique and error variances invariant across the African American and Caucasian student populations? (step 4)

5. Are the covariance matrices of subtest scores equivalent for African American and Caucasian students? (step 5)

Chapter 4

Results

The present study was undertaken to investigate the factor structure of the WISC III for a sample of African American and Caucasian students. The study explored the degree to which the WISC III measured the same underlying constructs for a sample of African American students as it did for a sample of Caucasian students. It was hypothesized that there would not be a statistically significant difference in the factor structure of the WISC III for the African American and Caucasian students. It also was hypothesized that there would be no statistically significant differences when comparing the covariance matrices, factor loadings, the factor correlation, factor variances and subtest unique and error variances in the WISC III for a sample of African American and Caucasian students.

The results of this study are presented in three sections. Section One describes the demographic characteristics of the sample. Section Two provides a description of the descriptive statistics and factor model of the two samples. Section Three presents the results of the multi-sample confirmatory factor analysis.

Section One

Participants. The initial sample for the study included 880 potential participants from the categories created to represent the African American and Caucasian groups based on ethnicity, gender, IQ category and age category. After central files records

were gathered, there were 545 (62% of the pool) participants included from the Hillsborough County Public School System who had been referred for additional testing for either superior or insufficient academic or behavioral progress. The intent of this study was to examine specific aspects of the factor structure of the WISC III in the African American and Caucasian populations. The WISC III Full Scale IQ and scale score data (Information, Similarities, Arithmetic, Vocabulary, Comprehension, Digit Span, Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding) for 291 Caucasian and 254 African American students were accessed. The sample consisted of 545 children, 264 male and 281 female, distributed across 8 age categories (6-6.11, 7-7.11, 8-8.11, 9-9.11, 10-10.11, 11-12.11, 13-13.11, 14-16.11). The age categorization was replicated from data reported in the WISC III technical manual (Wechsler, 1991). The sample also consisted of 7 Full Scale IQ categories (<55, 55-69,70-84, 85-99, 100-114, 115-129, >130). The IQ categorization was based on the standard deviation (SD=15) of WISC III Full Scale IQ scores. All participants were given the Wechsler Intelligence Scale for Children- Third Edition (WISC III). Table 1 details the breakdown of ethnicity and numbers of children at each age level included in this study. There were 4 persons without age level data that were not included in this table. Table 2 provides demographic data of the sample at each IQ category by ethnicity.

Table 1

Demographic Characteristics of the Sample: Participants by Ethnicity and Age Level

Age	Caucasian	African American	TOTAL
6 (6.0-6.11)	26	19	45
7 (7.0-7.11)	29	31	60
8 (8.0-8.11)	35	35	70
9 (9.0-9.11)	42	42	84
10 (10-10.11)	42	30	72
11-12 (11.0-12.11)	41	33	74
13 (13.0-13.11)	37	34	71
14-16 (14.0-16.11)	35	30	65
TOTAL	287	254	541

Table 2

Demographic Characteristics of the Sample: Participants by Ethnicity and IQ Level

IQ	Caucasian	African American	TOTAL
<55	19	32	51
55-69	50	36	86
70-84	41	46	87
85-99	53	47	100
100-114	55	42	97
115-129	39	32	71
>130	34	19	53
TOTAL	291	254	545

The group data were analyzed to ensure that the African American and Caucasian samples were comparable. With respect to race, 46.61% of the total sample consisted of African Americans while 53.39% were Caucasian. The gender composition of the total sample consisted of 51.56% males and 48.44% females. A chi square analysis was conducted to determine if there were statistically significant differences in the ethnic and gender balance of the sample. The results showed that the balance of genders for African Americans was not statistically significantly different from the balance of Caucasians. ($\chi^2(1)=.259$, $p=.611$) (see Table 3).

Table 3

Demographic Characteristics of the Sample: Participant Ethnic and Gender Chi Square Table

Frequency Percent	Female	Male	Total
African American	126	128	254
	23.12	23.49	46.61
Caucasian	138	153	291
	25.32	28.07	53.39
Total	264	281	545
	48.44	51.56	100.00

In terms of the age categories, there were no statistically significant differences found between the groups by race across the 8 age categorizations (6-6.11, 7-7.11, 8-8.11, 9-9.11, 10-10.11, 11-12.11, 13-13.11, 14-16.11) (see Table 4).

Table 4

Demographic Characteristics of the Sample: Participant Age Categorization Chi Square Table

Frequency Percent	1 (6-6.11)	2 (7-7.11)	3 (8-8.11)	4 (9-9.11)	5 (10-10.11)	6 (11-12.11)	7 (13-13.11)	8 (14-16.11)	Total
African	19	31	35	42	30	33	34	30	254
American	3.51	5.73	6.47	7.76	5.55	6.10	6.28	5.55	46.95
Caucasian	26	29	35	42	42	41	37	35	287
	4.81	5.36	6.47	7.76	7.76	7.58	6.84	6.47	53.05
Total	45	60	70	84	72	74	71	65	541
	8.32	11.09	12.94	15.53	13.31	13.68	13.12	12.01	100.0

Additionally, there were no statistically significant differences found between the two races in the distributions across the 7 IQ categories (<55, 55-69,70-84, 85-99, 100-114, 115-129, >130) ($\chi^2(6)=10.45$, $p=107$) (see Table 5).

Table 5

Demographic Characteristics of the Sample: Participant IQ Categorization Chi Square Table

Frequency Percent	<55	55-69	70-84	85-99	100-114	115-129	=/>130	TOTAL
African	32	36	46	47	42	32	19	254
American	5.87	6.61	8.44	8.62	7.71	5.87	3.49	46.61
Caucasian	19	50	41	53	55	39	34	291
	3.49	9.17	7.52	9.72	10.09	7.16	6.24	53.39
TOTAL	51	86	87	100	97	71	53	545
	9.36	15.78	15.96	18.35	17.80	13.03	9.72	100.00

Section Two

Descriptive Statistics. Univariate statistics were calculated to examine the normality of the distribution of scores by race across the 10 subtests of the WISC III (Picture Completion, Information, Coding, Similarities, Picture Arrangement, Arithmetic, Block Design, Vocabulary, Object Assembly, and Comprehension). Table 6 provides the descriptive statistics for each variable by race.

Table 6

Descriptive Statistics for the WISC III Subtests by Race

Variable Name	Race	Mean	Standard Deviation	Skewness	Kurtosis
Picture Completion	AA	8.09	4.12	-.03	-.63
	C	8.90	4.01	-.10	-.51
Information	AA	7.81	4.41	.34	-.88
	C	8.97	4.41	.06	-.96
Coding	AA	9.07	4.41	.15	-.78
	C	8.84	4.18	.10	-.79
Similarities	AA	8.22	4.62	.08	-1.02
	C	9.14	4.55	-.05	-.85
Picture Arrangement	AA	8.02	4.53	.30	-.57
	C	8.82	4.70	.02	-.91
Arithmetic	AA	8.18	4.25	.12	-.72
	C	8.45	4.28	-.01	-.73
Block Design	AA	8.14	4.31	.11	-.71
	C	9.22	4.80	.12	-.64
Vocabulary	AA	8.15	4.53	.27	-.75
	C	9.16	4.45	.08	-.72
Object Assembly	AA	8.00	4.06	.06	-.37
	C	8.86	4.19	.00	-.52
Comprehension	AA	8.65	4.67	.06	-.85
	C	9.49	4.80	-.14	-.90

Note. AA=African American, C=Caucasian

The mean scores ranged between 7.81 and 9.49 for the African American and Caucasian samples. The standard deviations for both groups fell between 4.01 and 4.80. The obtained mean scores for each subtest of the WISC III were lower than subtest scores reported in the WISC III manual (Weschler, 1991). In the WISC III manual, the average subtest mean score fell between 9.8 and 10.09. The average standard deviations as reported in the manual, fell between 2.5 and 3.8. The obtained subtest standard deviations for the current study were greater than those reported in the WISC III manual. Therefore, the sample average subtest scores and standard deviations represented more variability within the sample than that which was reported in the WISC III standardization sample.

When examining the variability of the sample, there were no notable degrees of skewness found in these two groups. However, there were consistently negative kurtosis values, which indicated that the scores tended to fall in close proximity of each other. Meaning, there were less outliers than what would be expected in a normal distribution. There was a departure from normality as evidenced in the negative kurtosis values. The majority of the kurtosis values fell between -.5 and -1.0. Therefore, the distributions of both samples were relatively platykurtic. While the skewness and kurtosis values obtained reflect a departure from what would be expected in the normal distribution, the departures from normality were not severe enough to warrant normality corrections.

Based on the demographic characteristics and preliminary statistical analysis of the sample, there appeared to be sufficient evidence to support an investigation of the proposed research questions using maximum likelihood estimation in the confirmatory factor analysis. Although there were data that violated the normal distribution

assumption, the relatively small departures from normality were not significant enough to suggest a different data analysis process.

Additionally, the correlation matrices for the African American and Caucasian groups are presented in Tables 7 and 8. There were notable similarities across the groups. The obtained intercorrelations ranged from .55 to .88. The largest discrepancy was found between the interrelation of Picture Completion and Coding for the African American (r=.65) and Caucasian (r=.56) groups with a difference of .09. The intercorrelations for the standardization sample are reported in the WISC III manual (Weschler, 1991). However, the intercorrelations reported in the manual were lower than those obtained with the current study. For example, the intercorrelations between the Information and Similarities subtests for the African American and Caucasian groups were .84 and .85, respectively. In the WISC III manual the average intercorrelation between the Information and Similarities subtest across all age groups was .66. The intercorrelations reported in the manual averaged between .24 and .70.

TABLE 7

Correlation Matrix for the African American Sample

	PIC	INF	COD	SIM	PICA	ARI	BLO	VOC	OBJ	COM
PIC	1.00000	0.73172	0.65022	0.71109	0.75015	0.73339	0.75314	0.74333	0.70790	0.70771
INF	0.73172	1.00000	0.66671	0.84375	0.72318	0.78318	0.72173	0.85724	0.70054	0.82343
COD	0.65022	0.66671	1.00000	0.65764	0.67295	0.71253	0.71217	0.68089	0.63210	0.65260
SIM	0.71109	0.84375	0.65764	1.00000	0.75086	0.77273	0.70466	0.83322	0.68627	0.83162
PICA	0.75015	0.72318	0.67295	0.75086	1.00000	0.75047	0.75369	0.74562	0.68335	0.72314
ARI	0.73339	0.78318	0.71253	0.77273	0.75047	1.00000	0.75245	0.78362	0.68303	0.75056
BLO	0.75314	0.72173	0.71217	0.70466	0.75369	0.75245	1.00000	0.73696	0.75579	0.66777
VOC	0.74333	0.85724	0.68089	0.83322	0.74562	0.78362	0.73696	1.00000	0.68318	0.81871
OBJ	0.70790	0.70054	0.63210	0.68627	0.68335	0.68303	0.75579	0.68318	1.00000	0.65356
COM	0.70771	0.82343	0.65260	0.83162	0.72314	0.75056	0.66777	0.81871	0.65356	1.00000

Note. PIC=Picture Completion, INF=Information, COD=Coding, SIM=Similarities, PICA=Picture Arrangement, ARI=Arithmetic, BLO=Block Design, VOC= Vocabulary, OBJ=Object Assembly, COM=Comprehension

TABLE 8

Correlation Matrix for the Caucasian Sample

	PIC	INF	COD	SIM	PICA	ARI	BLO	VOC	OBJ	COM
PIC	1.00000	0.73820	0.55607	0.71057	0.70251	0.67943	0.76598	0.71969	0.72193	0.68998
INF	0.73820	1.00000	0.59527	0.84542	0.74787	0.80047	0.71824	0.88493	0.67178	0.82693
COD	0.55607	0.59527	1.00000	0.60093	0.69718	0.69719	0.67087	0.62852	0.54927	0.62344
SIM	0.71057	0.84542	0.60093	1.00000	0.70929	0.75175	0.67902	0.84833	0.61389	0.83522
PICA	0.70251	0.74787	0.69718	0.70929	1.00000	0.75010	0.76944	0.73379	0.71104	0.73214
ARI	0.67943	0.80047	0.69719	0.75175	0.75010	1.00000	0.73548	0.79583	0.65727	0.76796
BLO	0.76598	0.71824	0.67087	0.67902	0.76944	0.73548	1.00000	0.71760	0.79664	0.68290
VOC	0.71969	0.88493	0.62852	0.84833	0.73379	0.79583	0.71760	1.00000	0.65209	0.85925
OBJ	0.72193	0.67178	0.54927	0.61389	0.71104	0.65727	0.79664	0.65209	1.00000	0.66107
COM	0.68998	0.82693	0.62344	0.83522	0.73214	0.76796	0.68290	0.85925	0.66107	1.00000

Note. PIC=Picture Completion, INF=Information, COD=Coding, SIM=Similarities, PICA=Picture Arrangement, ARI=Arithmetic, BLO=Block Design, VOC= Vocabulary, OBJ=Object Assembly, COM=Comprehension

Factor Model. The first step began with a comparison of the two samples to examine model fit. All parameter estimates of the factor solution were allowed to vary. This first step was equivalent to conducting separate confirmatory factor analyses for the Caucasian and African American participants. Figure 2 exhibits the parameter estimates for the African American sample for the specified model. The results of this analysis yielded a statistically significant chi square ($\chi^2(34)=59.75$, p=.004). The statistically significant chi square suggests that the specified model is not a perfect model for the data. It has been noted that the chi square can be statistically significant when a model is good, but not perfect. If forced to the specified model, the data yielded a statistically significant Bentler-Bonnett Nonnormed Fit Index (NNFI) of .987 and a Bentler's Comparative Fit Index (CFI) of .990. These fit statistics reveal that when the data were forced to fit to the specified model, there was a relatively good fit.

Similar results were found when the independent confirmatory factor analysis was conducted with the Caucasian group. Figure 3 exhibits the parameter estimates for the Caucasian sample for the specified model. A larger chi square value was obtained with this group ($\chi^2(34)=131.94$, p=.0001). Again suggesting that the specified model was not the ideal model fit for the data. However, the NNFI of .959 and CFI of .969 suggest that when the data were forced to the specified model, there was an acceptable fit.

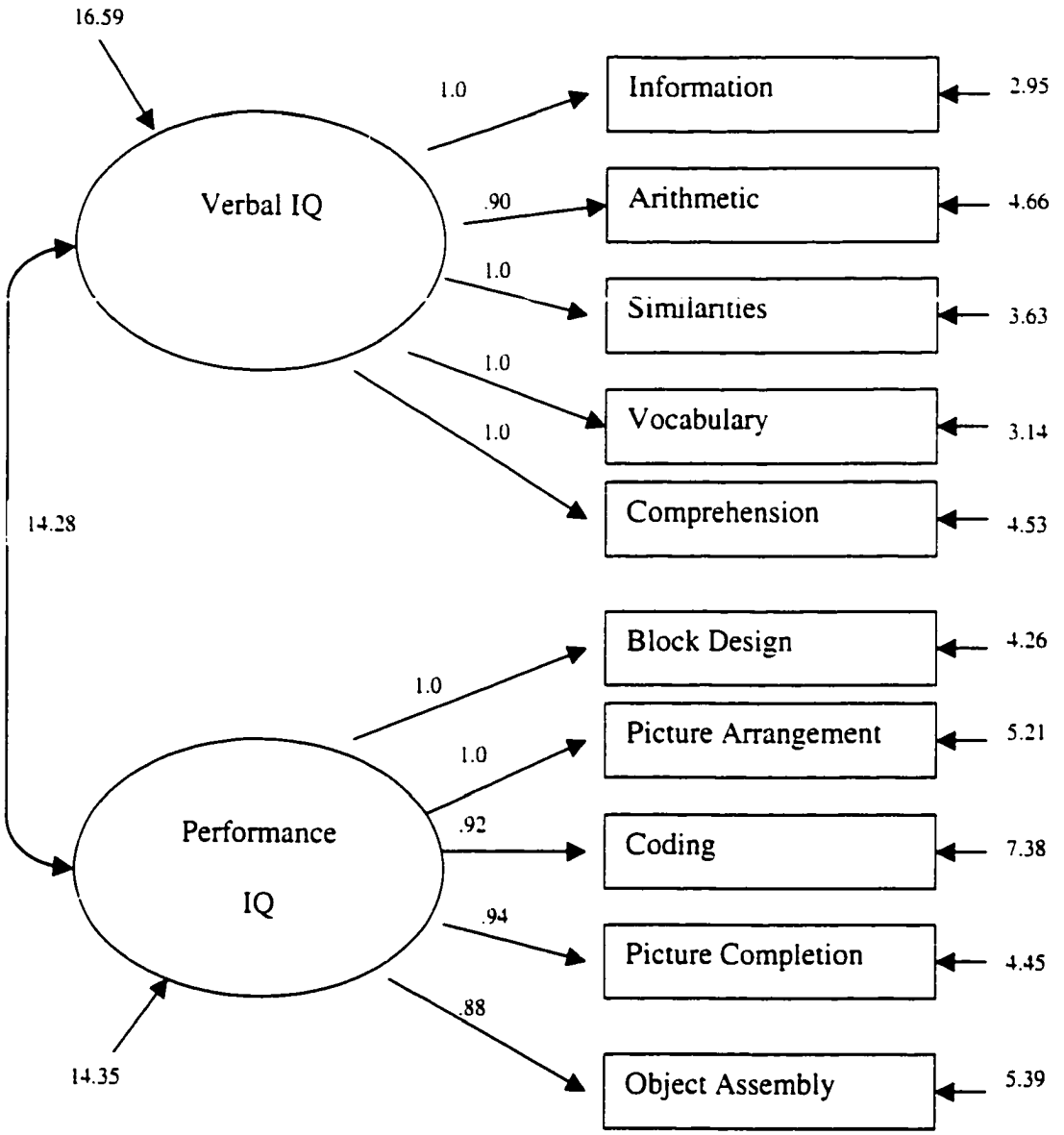


Figure 2. WISC III Factor Model for the African American Sample.

Note. All parameter values are statistically significantly different from zero ($p<.05$).

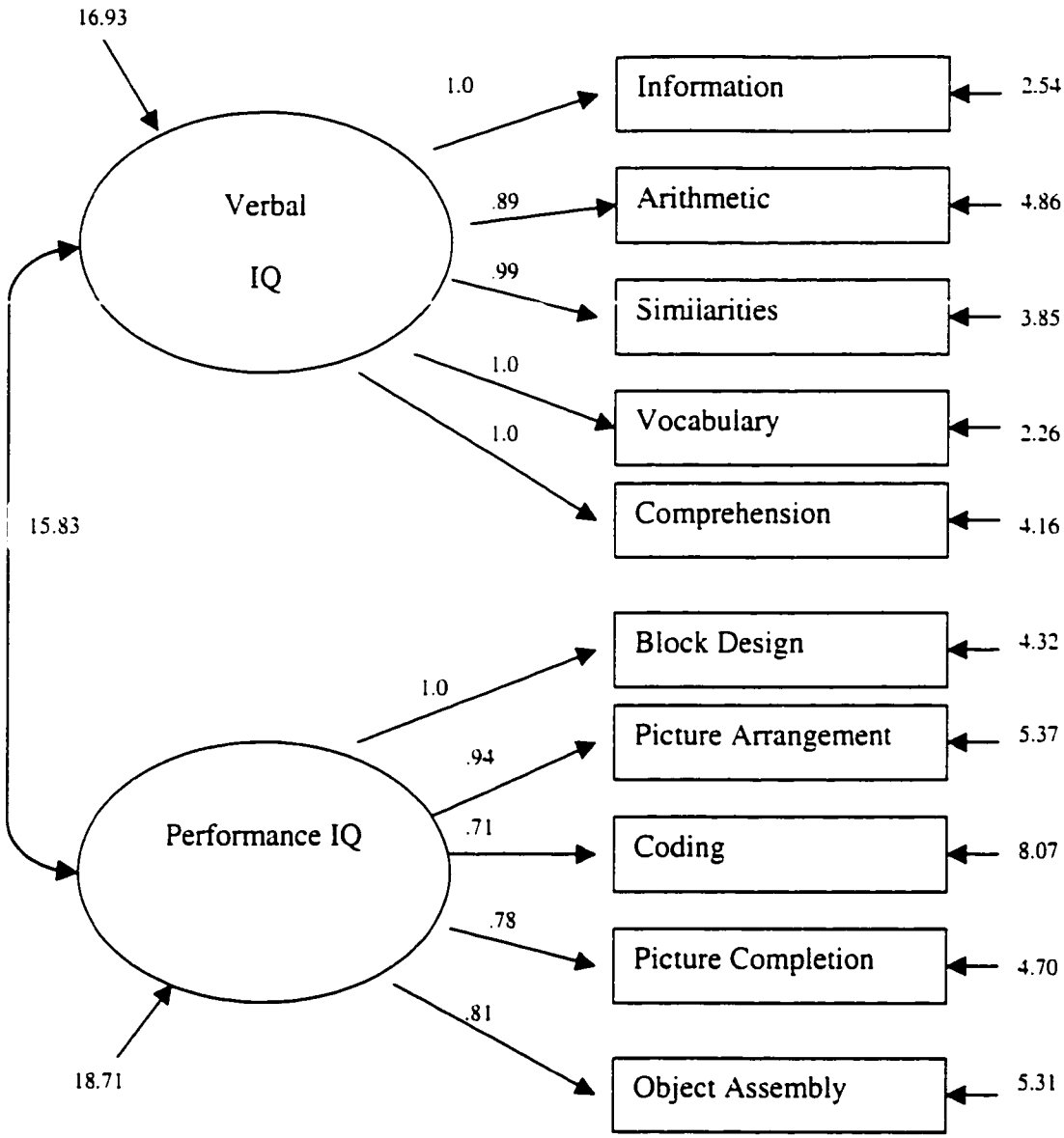


Figure 3. WISC III Factor Model for the Caucasian Sample.

Note. All parameter values are statistically significantly different from zero ($p<.05$).

Next, the two groups were analyzed using a multi-sample confirmatory factor analysis. The results of the analysis of the two groups fit to the two-factor WISC III model yielded a statistically significant chi square value ($\chi^2(68)=191.70$, $p<.001$). As previously stated, the significant chi square values obtained show that the specified two-factor WISC III

model does not completely explain the observed covariances. It has been noted that the chi square statistic is dependent on sample size (Keith et al., 1995), therefore, additional fit statistics are reported. The Comparative Fit Index (CFI) of .979 and Bentler-Bonett Nonnormed Fit Index (NNFI) of .972 suggest a good fit of the data to the model. These fit statistics provide additional information regarding the fit of the data to the model if the data were forced to fit the hypothesized model. When forcing the data to fit the hypothesized model, the fit statistics examine whether the model represents the data well without depending upon sample size. The additional fit statistics show that the WISC III model provides a relatively good fit to the data for both groups.

Section Three

Multi-sample Confirmatory Factor Analysis (MCFA). Since the model appears to be a reasonable approximation for each group, the multi-sample confirmatory factor analysis was conducted (MCFA). As previously stated, the EQS statistical data analysis program (Bentler, 1995), was used to conduct the MCFA. The step procedures outline model fit in providing the most parsimonious explanation of test performance across groups. Table 9 outlines the results of the 4-step MCFA process.

Table 9

Comparison of the Fit of Model Factors Specifying Various Degrees of Freedom across Ethnic Groups

Model Description	χ^2 (df)	p	CFI	NNFI	χ^2 change (df)	p
Step 1: Independent group analysis of data fit to the model	197.92 (68)	.005	.968	.959		
Step 2: Factor loadings are invariant	204.20 (75)	<.001	.978	.973	6.28 (8)	>.05
Step 3: Factor loadings, the factor correlation and factor variances are invariant	207.94 (78)	<.001	.977	.974	3.74 (3)	>.05
Step 4: Factor loadings, the factor correlation, factor variances, error & unique variances are invariant	214.17 (88)	<.001	.978	.978	6.23 (10)	>.05

In step 2 the factor loadings were held constant for both groups. The results show that there was not a statistically significant change in chi square obtained by making the factor loadings the same for both the African American and Caucasian samples ($\Delta\chi^2(7)=6.28$, $p>.05$). The measures of goodness of fit CFI (.978) and NNFI (.973) showed that with these constraints, the specified model was still a good fit for both the African American and Caucasian samples.

In step 3, the factor loadings, factor variances and the factor correlation were held constant across groups. There was not a statistically significant change in the fit from the previous step ($\Delta\chi^2(3)=3.74$, $p>.05$). The fit indices (CFI=.98 and NNFI=.97) confirm that when the factor loadings, factor variances and the factor correlation are held constant the model still fits the data well.

For step 4, all parameter estimates were constrained to be equal (factor loadings, the factor correlation, factor variances, and unique and error variances). The results obtained showed that when all parameter estimates were constrained, there was not a statistically significant change in the chi square ($\Delta\chi^2(10)=6.23$, $p>.05$). Therefore, when all parameter estimates were invariant, the specified model of the WISC III was a good fit for both the African American and Caucasian samples (CFI=.98, NNFI=.98). Because there were not statistically significant differences in the model for both groups, it appears that the same model could be used for both groups. This model is represented with the parameter estimates in Figure 4.

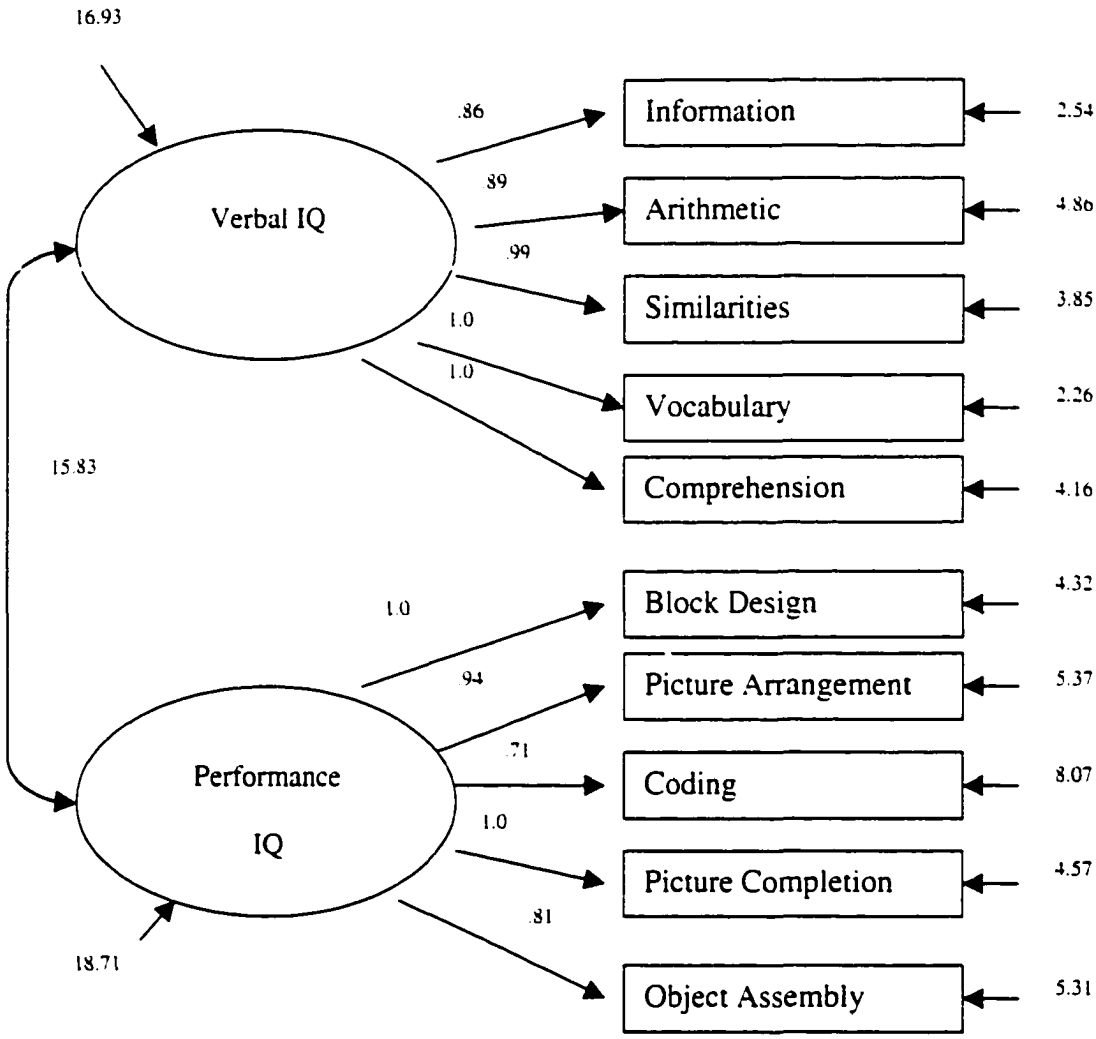


Figure 4. WISC III Factor Model for the African American and Caucasian Samples.

With the obtained results, it appears that the specified model provides a relatively good fit to the current study data. In order to provide more interpretable results, Figure 5 represents the standardized solution for the data from both groups. It appears that the subtests representing the Verbal IQ factor have a higher correlation with the factor than

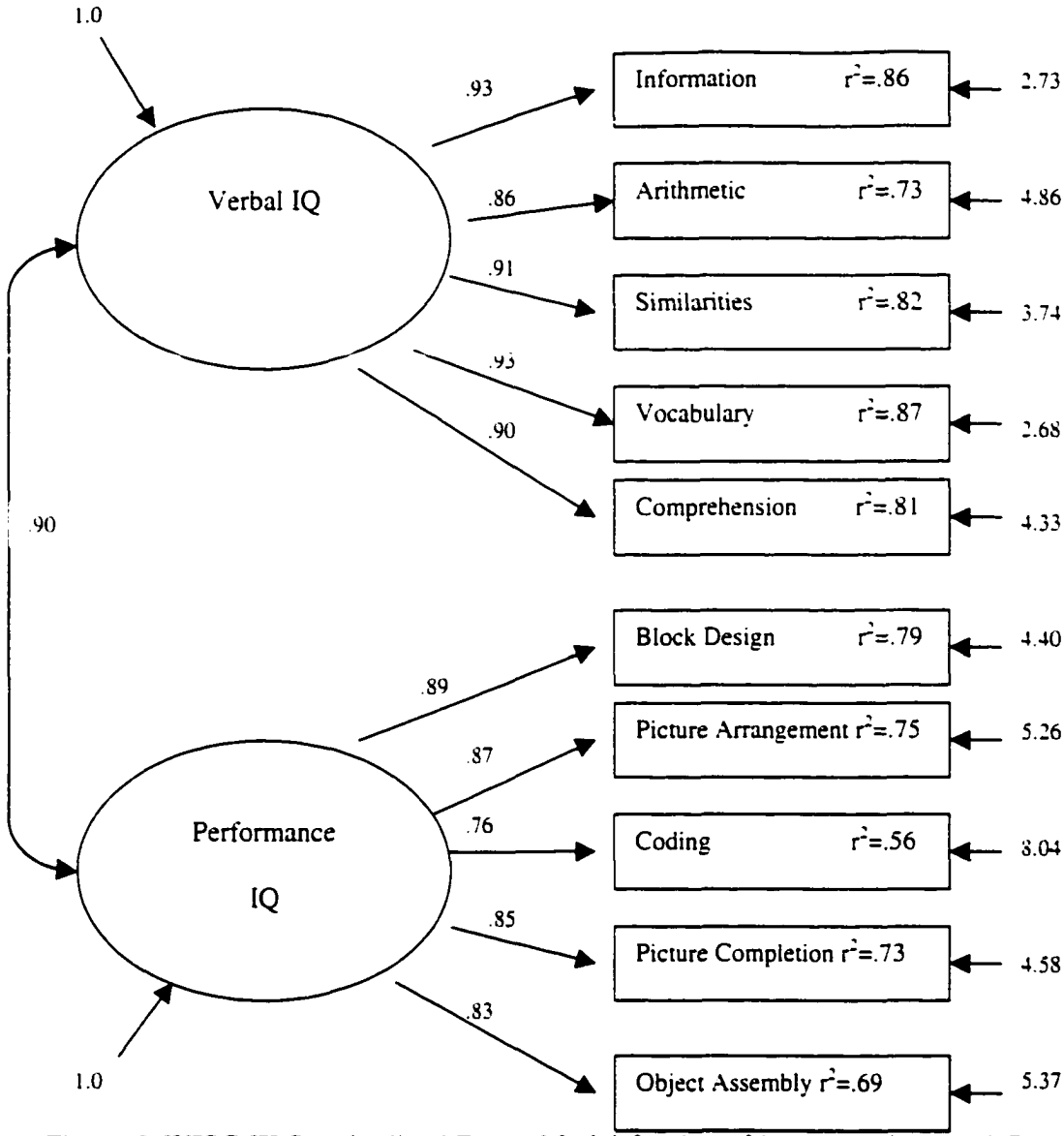


Figure 5. WISC III Standardized Factor Model for the African American and Caucasian Samples.

those that are representative of the Performance IQ factor. The path coefficients for the Verbal IQ factor range from .86 to .93 while the path coefficients for the Performance IQ factor range from .76 to .89. The r^2 are reported for each subtest. The obtained r^2 shows the amount of variability within a specific subtest that can be accounted for by the factor. When examining the subtests that load on the Verbal IQ factor, the Verbal factor

accounted for 87% of the variability in the Vocabulary subtest and 73% of the Arithmetic subtest. The Performance IQ factor accounted for 79% of the Block Design subtest and 56% of the Coding subtest.

Steps 1 through 4 showed similarity in parameter values given a two-factor model. A more general test can be done by testing the equivalence of the variance/ covariance matrices. This analysis assisted in determining whether the WISC III measures the same constructs or attributes across both groups (Witta & Keith, 1997). The results yielded a chi square of 58.82, which was not statistically significant (p=.38).

In summary, when the two groups were constrained to fit the specified two-factor model of the WISC III, the data obtained revealed that the two-factor model may not have been a perfect fit, but did provide an adequate representation of the data. There were no significant differences found when using confirmatory factor analysis to investigate ethnic differences in the model fit when constraining various parameters.

Chapter 5

Discussion

The purpose of this study was to investigate the factor structure of the WISC III for a sample of referred African American and Caucasian students. The participants consisted of 545 children between the ages of 6 and 16 who had been referred for either superior or insufficient academic or behavioral achievement between the years 1995-1997 in the Hillsborough County Public School System. These students were all given the Weschler Intelligence Scale for Children- Third Edition. This chapter discusses the hypothesized research questions regarding the WISC III factor structure. Secondly, the obtained results are related with previous research on the WISC III factor structure followed by the limitations of the current research. The chapter concludes with a discussion of future areas of research.

Response to Research Questions

The first research question examined whether there would be statistically significant differences when comparing construct bias in the WISC III for a sample of African American and Caucasian students. From the results obtained, there were no statistically significant differences in the parameter estimates of African American and Caucasian students who had been referred for additional testing using the WISC III. Therefore, we failed to reject the null hypothesis that there were no statistically significant differences in the factor structure of the two factor model of the WISC III for

African American and Caucasian students. The second research question examined whether there would be statistically significant differences in the factor loadings of the WISC III when comparing a sample of African American and Caucasian students. The results obtained revealed a good fit of the data to the WISC III model when factor loadings were invariant (CFI=.978, NNFI=.973). Therefore, it appeared that the factor structure for the WISC III was still a good fit for the African American and Caucasian groups. The third research question asked if there would be a statistically significant difference in the factor variances and the factor correlation of the WISC III when comparing a sample of African American and Caucasian students. When the researcher held the factor loadings, factor variances and the factor correlation constant (invariant), the results obtained did not reveal a significant change in the obtained chi square ($\Delta\chi^2$=3.74, p>.05). The WISC III model was a good representation of the African American and Caucasian student data (CFI=.977, NNFI=.974).

The fourth research question examined whether there would be statistically significant differences in the subtest unique and error variances of the WISC III when comparing a sample of African American and Caucasian students. The results obtained from the current study revealed that there were no statistically significant differences in the obtained chi square when the factor loadings, factor variances, the factor correlation, and subtest unique and error variances were invariant ($\Delta\chi^2$=6.23, p>.05). In fact, the model fit to the data for both the African American and Caucasian groups improved when these parameter estimates were held constant (CFI=.978, NNFI=.978). Therefore, we failed to reject the null hypothesis that their would be no statistically significant differences found between the African American and Caucasian groups when all

parameter estimates were held constant. The final research question examined whether there would be statistically significant differences in the covariance matrices of subtest scores on the WISC III for a sample of African American and Caucasian students. The results of the current study revealed that there were no statistically significant differences in the covariance matrices for the African American and Caucasian students using the prescribed WISC III model (χ^2=58.82, p>.05). In conclusion, there were no statistically significant differences found between the sample of African American and Caucasian students in model fit for the WISC III.

Relationship to Previous Research

Early exploratory and confirmatory factor analytic studies provided contravening results regarding the construct validity of the WISC III. Slate and Jones (1995) found subtests that did not load appropriately on the Verbal and Performance IQ scales when comparing groups of African American and Caucasian students. However, this study consisted of a sample size of 58 students, which hinders the generalization of the results. While the Slate and Jones (1995) study provided information regarding the subtest relationships with the Verbal and Performance factors of the WISC III, it did not investigate the goodness of the model fit to the data. In contrast, Kush and Watkins (1997) found in their study that the data could be best represented by two factors, traditionally labeled, Verbal and Performance. Together, these two factors accounted for 57% of the total test variance, a number greater than what was reported in the test manual (43%) for these two factors. In fact, the authors found a high degree of factorial similarity between their sample and the standardization sample for both these factors with coefficients of congruency equaling .99 for the Verbal factor and .98 for the Performance

factor. While exploratory factor analysis is a good method for identifying latent constructs that account for intercorrelations among a set of variables, it produces mathematically indeterminate results (Gorsuch, 1983). Therefore, the results of this study are not directly comparable with the current study's results. However, Kush and Watkins did investigate the validity of the WISC III two-factor model. The current research findings suggest a good fit of their prescribed model for a large sample of African American and Caucasian students. The current data gathered supports the research conducted by Kush and Watkins demonstrating a good model fit for African American students.

The present findings are consistent with previous research findings indicating that there were no statistically significant differences found in the factor structure of the WISC III between African American and Caucasian samples of students (Kush & Watkins, 1997; Kush, Watkins, Ward, Ward, Canivez & Worrel, 1999; Vance & Engin, 1978). Kush, Watkins, Ward, Ward, Canivez and Worrell (1999) conducted exploratory and confirmatory factor analyses using the WISC III standardization sample to compare the performance African American and Caucasian students. Using the 12 subtests of the instrument, the authors found good to excellent factorial similarity between the African American and Caucasian students for the two-factor model (Verbal and Performance). Upon further investigation using confirmatory factor analysis techniques, the authors found that the best overall fit of the data for both groups fell with a four-factor model. The authors conclude that the WISC III provides a good general interpretation of the general intelligence factor, *g*. They suggest that any interpretation beyond that be done using the four-factor as opposed to the two-factor model. Again, the results of this study

confirm the current research findings that the two-factor model is a good fit for both the sample African American and Caucasian student population. However, the Kush et al. study does lead to future research questions regarding whether the best fit of the data could have been determine from a four-factor model.

Referral Bias & Placement Disproportionality

From the results obtained, it appears that the WISC III is a fairly adequate instrument when investigating measurement invariance for a sample of referred African American and Caucasian students. Yet, the fact remains that a disproportionate number of African Americans are represented in specific special education categories. In Hillsborough County, were the study was conducted, the school district appeared to have an overrepresentation of African American student in the total special education population. Bias can occur in a number of areas within the process and procedures of student assessment. The bias could be in the test (Gould, 1981; Medina & Neill, 1990), in the testing procedures (Fuchs & Fuchs, 1989a) or in the classroom teachers who are referring children for testing (Algozzine, Christenson, & Ysseldyke, 1982; Algozzine, Mercer, & Countermine, 1977; Bahr, Fuchs, Stecker, & Fuchs, 1991; Ysseldyke, Algozzine, Regan & McGue, 1981).

Various researchers have investigated the referral bias that may be present with the classroom teacher and other decision makers (Algozzine, 1977; Egeland, & Abery, 1991; Kauffman, Wong, Lloyd, Hung, & Pullen, 1991; Thurlow & Ysseldyke, 1982; Ysseldyke & Algozzine, 1981). Ysseldyke et al. (1981) found that when presented with certain variables (the reason for referral, socioeconomic status, physical attractiveness, IQ scores, academic achievement scores, perceptual motor scores, personality scores,

language test scores, adaptive behavior scale scores, behavioral observation results and behavioral checklists), decisions regarding special education referrals most frequently were made based on the reason for referral. Meaning, all students have a greater chance of being placed in special education programs based mostly on their teacher's opinion of the presenting problem. Additionally, decision makers used information regarding high socioeconomic status, physical attractiveness, intellectual and academic achievement to aid in decisions regarding special education placements. In looking further at these issues, African American students are hindered by the extraneous variables that impact decision makers. With regard to high socioeconomic status serving as a buffer, the majority of African American students fall within the middle to lower class and are not benefited by high socioeconomic status. When examining the issue of physical attractiveness, the traditional African facial features are not commonly referred to as attractive by the majority culture. Studies by Andrews, Wisniewski, and Mulick (1997) also investigated the impact of physical characteristics in the referral process. These authors found that when the influence of age, height, and weight were taken into account, teachers were more likely to refer a child for special education when their height and weight were perceived to be above average for their age. Bias could also occur based on within person factors. When examining within teacher variables, Soodak and Podell (1993) found that when teachers had high personal self-efficacy they tended to think that the general education setting was appropriate for children with low socioeconomic status and the converse also was true.

These studies provide evidence that there may be other reasons for differences that could explain the disproportionate number of African American students present in

certain special education categories. It is noted that there are a disproportionate number of African American students enrolled in the educable mentally handicapped special education program. One hypothesis for the continued disproportionality could be that without the presence of referral bias, the ethnic differences in students referred for special education may not be as great as they currently are. Meaning, if a more structured and objective process was instituted, personal characteristics such as socioeconomic status, height, weight and opinions about the referral problem would not greatly influence the referral process. Most of the factors that researchers found to greatly influence referral decisions were things that were out of the control of the child. Future research should focus on referral bias and the process by which students are identified for special education. It is necessary to continue to investigate means by which children are assessed fairly, based on their educational performance and not on inherent, physical characteristics.

Because school psychologists are aware of these barriers that prevent an accurate assessment of students' needs, they are the appropriate persons to being the process of educating teachers and school staff in how to make more objective educational decisions. The school psychologist should have a more prominent role in determining the appropriateness of the referral for special education evaluation through assisting school personnel in making the most non-biased decision process possible.

Limitations

Several limitations were evident in this research. First, the WISC III protocol results were accessed by examining participant archival records. This served as a limitation for two reasons. First, the researcher lacked the ability to manipulate the school

and assessment conditions. In the study, all students in the Hillsborough County Public School System did not have an equal chance of being a participant. The researcher had no control of which participants were tested and which scores reported to Hillsborough County School System central files. Second, fully random assignment to conditions was not possible.

Another threat to the internal validity of this study may have been selection-maturation issues. The effects of their particular school and school personnel may have impacted the participants selected. However, the participants were randomly selected from a test pool of students throughout Hillsborough County. Therefore any impact based on school particularities should have been negated by randomized school representation. Another threat to the internal validity of the study was instrumentation. Because the WISC III is individually administered and scored, the scoring process used may vary between school psychologists. Yet, this threat was limited by the fact that the WISC III was administered by state certified school psychologists. Each of these psychologists has received graduate level advanced training in the WISC III administration procedures. Their training included frequent observations of testing procedures by highly trained university and field supervisors. In conclusion, the school psychologists followed a prescribed, standardized set of procedures to administer the WISC III.

The generalizability of this experiment to the population was restricted for several reasons. First, a convenience sampling procedure was used to obtain participants. The participants were chosen based on ease of entry and acceptability of the researcher as part of the school environment. Additionally, as a control for investigator bias, the researcher had volunteers to assist in data collection from the district central files and in coding the

participant names and ethnicity to a numerical format. However, the use of the volunteers to collect data may have served an additional limitation. While these persons were trained and obtained interrater agreement prior to data transfer, they may have collect the student data that were easy to obtain. Therefore, not all identified students had an equal chance of being selected. Additionally, only students from the Hillsborough County Public School System were chosen to participate, which may inhibit generalization of the results to other children across the country. Finally, the participants selected may not represent the total school population.

Recommendations for Future Research

With the aforementioned issues in mind, there are several areas that should be investigated to work towards a solution in the disproportionate placement of minority students in handicapping special education conditions. First, school psychologists, who frequently administer tests such as the WISC III could investigate other assessment methods. Current research has investigated a dynamic assessment process during which the school psychologist uses methods such as parent, teacher and student interviews and curriculum based assessment to determine whether a child is meeting grade level expectations in the curriculum. One argument against using standardized tests such as the WISC III has been that the test is bias because it does not take into consideration that children have differing live and educational experiences that cannot be measured using a test created by one group of researchers. Therefore it is recommended that students be evaluated using the curriculum in which they receive instruction. Curriculum based assessment provides the teacher with direct information regarding the students' progress in the identified curriculum. Monitoring student progress using the students' curriculum

would provide a more direct assessment of the student's skill acquisition using the curriculum that they are being taught in. School psychologists could introduce such innovations in assessment to the public schools through training the teachers and staff that there are alternative assessment methods.

Future research should continue to examine the possible differential factor structure of the WISC III across various ethnic groups. Specifically, research should continue to investigate the performance of a large-scale general population of African American students on this instrument. Research to date has focused on the performance of students who have been either referred to or classified in a special education placement. Thereby limiting the generalization of the results to the larger African American population. Studies should be conducted to determine if the WISC III two-factor model provides good data fit for the general African American student population.

Additionally, future research areas include an investigation of which WISC III model is most appropriate for use based on theories of intelligence. Previous research has found validity in the three and four-factor model for the WISC III. While substantial data exist to validate the two-factor model, there is competing research that strongly validates the three-factor model (Carroll, 1993a, Carroll, 1993b, Kaufman, 1993, Keith & Witta, 1997). While the current research is investigating the validity of a three-factor model, future research should begin to explore the performance of various ethnic groups given the three-factor model to determine a model of "best fit." An increased knowledge of the interrelationships among these factors will be critical for psychologists who work with ethnically and linguistically diverse populations.

Finally, future research would be enhanced by continuing to examine the predictive validity of this instrument in forecasting academic achievement across majority and minority populations. Previous studies have provided contradicting evidence of the contribution of intelligence tests in the prediction of academic achievement (Gresham & Witt, 1997; Kaufman, 1994). Some studies have found that intelligence tests appear to be weak predictors of reading acquisition (Siegel, 1989,1992; Share, McGee & Silva, 1989, 1991). Additionally, many researchers argue that there is a weak link between intelligence tests and achievement because of the lack of an aptitude by treatment interaction (Ayers & Cooley, 1986; Ayers, Cooley, & Severson, 1988; Das, 1995; Das, Naglieri, & Kirby, 1995; Good, Vollmer, Creek, Katz, & Chowder, 1993). An aptitude by treatment interaction is the belief that the measurement of aptitudes (individual characteristics or traits) can predict the probability of success when given certain treatments (educational programs). It is the belief of some in the African American community that tests such as the WISC III do not and will not provide an adequate representation of the intellectual capabilities and academic potential of African American students. Given the lack of an aptitude by treatment interaction, the concern is that there must be other assessment methods that are more reflective of student cognitive and academic functioning than the WISC III. Decisions regarding special education placement could be challenged because the assessment measures that are used have not been found to predict student success in the educational program. However, there are numerous variables that may be considered (e.g., teacher efficacy, individualized curriculum) when determining program efficacy. Yet there still is an undeniable relationship between performance on intellectual measures and performance on academic

achievement measures (Glutting, Kelly, Boehm & Burnett, 1989). Therefore, future researchers should investigate the relationship of the WISC III with academic achievement in order to provide preventative education programming thereby minimizing academic underachievement.

REFERENCES

Algozzine, B. (1977). Perceived attractiveness and classroom interactions. Journal of Experimental Education, 46, 63-66.

Algozzine, B., Christenson, S., & Ysseldyke, J. E. (1982). Probabilities associated with the referral to placement process. Teacher Education and Special Education, 5, 19-23.

Algozzine, B., Mercer, C., & Countermine, T. (1977). Labeling exceptional children: An analysis of expectations. Exceptional Children, 44, 131-132.

Algozzine, B., Ysseldyke, J. E., & Christenson, S. (1983). An analysis of the incidence of special class placement: The masses are burgeoning. Journal of Special Education, 17, 141-147.

American Educational Research Association, American Psychological Association, & The National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Anastasi, A. (1988). Psychological testing. New York: Macmillan.

Andrews, T. J., Wisniewski, J. J., & Mulick, J.A. (1997). Variables influencing teachers' decisions to refer children for school psychological assessment services. Psychology in the Schools, 34 (3), 239-244.

Arter, J. A., & Jenkins, J. R. (1979). Differential diagnosis-prescriptive teaching: A critical appraisal. Review of Educational Research, 49, 517-555.

Artiles, A. J., & Trent, S. C. (1994). Overrepresentation of minority students in special education: A continuing debate. Journal of Special Education, 27 (4), 410-437.

Ayers, R. R., Cooley, E. J. (1986). Sequential versus simultaneous processing on the K-ABC: Validity in predicting learning success. Journal of Psychoeducational Assessment, 4, 211-220.

Ayers, R. R., Cooley, E. J., & Severson, H. H. (1988). Educational translation of the Kaufman Assessment Battery for Children: A construct validity study. Journal of Psychoeducational Assessment, 4, 113-124.

Bahr, M. W., Fuchs, D., Stecker, P. M., & Fuchs, L. S. (1991). Are teachers' perceptions of difficult-to-teach students racially biased? School Psychology Review 20 (4), 599-608.

Barona, A., & Faykus, S. (1992). Differential effects of sociocultural variables on special education eligibility categories. Psychology in the Schools 29, 313-320.

Bersoff, D. N., & Hofer, P. T. (1990). The legal regulation of school psychology. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology (pp. 937-961). NY: Wiley.

Bollen, K. A. (1989). Structural equations with latent variables. New York: John Wiley & Sons.

Bracken, B. A. (Ed.) (1993). Monograph series advances in psychoeducational assessment: Wechsler Intelligence Scale for Children: Third Edition; Journal of Psychoeducational Assessment. Brandon, VT: Clinical Psychology Publishing.

Braden, J. P. (1995). The WISC III. The Twelfth Mental Measurements Yearbook. Lincoln, NE: University of Nebraska Press.

Braum, C., Neilson, A., & Dykstra, R. (1975). Teacher's expectations: Prime mover or inhibitor? Elementary School Journal, 181-188.

Browne, M. W. (1990). MUTMUM PC: User's guide. Columbus, OH: Ohio State University Press.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin, 105 (3), 456-466.

California Association of School Psychologists v. California Department of Education, C-93-2891 DLJ (N. D. Cal. 1994). [Electronic data tape]. (1994). Abstract from: Lexis-Nexis Academic Universe.

Carvajal, H. H., Hayes, J. E., Lackey, K. L., Rathke, M. L., Wiebe, D. A., & Weaver, K. A. (1993). Correlations between scores on the Wechsler Intelligence Scale for Children III and the General Purpose Abbreviated Battery of the Stanford Binet IV. Psychological Reports, 72 (3), 1167-1170.

Cazden, C. B., & Mehan, H. (1989). Principles for sociology and anthropology: Context, code, classroom and culture. In M. C. Reynolds (Ed.), Knowledge base for the beginning teacher (p. 47-57). Oxford, England: Pergamon Press.

Center for the Study of Social Policy. (1996). Kids count data book. Washington, DC: Author.

Chinn, P. C., & Hughes, S. (1987). Representation of minority students in special education classes. Remedial and Special Education, 8 (4), 41-46.

Cooper, H. M. (1979). Pygmalion grows up: A model for teacher expectation communication and performance influence. Review of Educational Research, 49 (3), 95-103.

Crawford v. Honig, C-89-0014 RFP, No. C-71-2270 RFP (N.D. Cal. 1992). [Electronic tape]. (1992). Abstract from: Lexis-Nexis Academic Universe.

Curtis, M. J., Hunley, S. A., Walker, K. J., & Baker, A. C. (1999). Demographic characteristics and professional practices in school psychology. School Psychology Review, 28 (1), 104-115.

D'Amato, R. C. (1990). A neuropsychological approach to school psychology. School Psychology Quarterly, 5,141-160.

Das, J. P. (1995).Neurocognitive approach to remediation: The PREP model. Canadian Journal of School Psychology, 9, 157-173.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1995). Assessment of cognitive processes. Needham, MA: Allyn & Bacon.

Doll, B., & Boren, R. (1993). Performance of severely language impaired students on the WISC III, language scales, and academic achievement measures. In B. Bracken (Ed.) Monograph series advances in psychoeducational assessment: Wechsler Intelligence Scale for Children: Third Edition; Journal of Psychoeducational Assessment. Brandon, VT: Clinical Psychology Publishing.

"Does race really make a difference in intelligence?" (1956, October 26). U.S. News and World Report, pp. 74.

Dunn, L. M. (1968). Special education for the mildly retarded: Is much of it justifiable? Exceptional Children, 23, 5-21.

Dusek, J. B. (1975). Do teachers bias children's learning. Review of Educational Research, 45 (4), 230-238.

Dwyer, K. & Stanhope, V. (1998). IDEA '97: Synopsis & recommendations. [On-line], National Association of School Psychologists. Available: http:\\www.naspweb.org.

Edelman, M. W., & Howe, H. (1985). Barriers to excellence: Our children at risk. Boston: National Coalition of Advocates for Students.

Eder, D. (1981). Ability grouping as a self fulfilling prophecy: A microanalysis of teacher student interaction. Sociology of Education, 54, 151-162.

Education for All Handicapped Children Act, 20 U. S. C. & 1401-1461 (1975).

Egeland, B., & Abery, B. (1991). A longitudinal study of high-risk children: Educational outcomes. International Journal of Disability, 38 (3), 271-287.

Elliott, S. N., & Boeve, K. (1987). Stability of WISC-R IQs: An investigation of ethnic differences over time. Educational and Psychological Measurement, 47, 461-465.

Entwisle, D. R., & Alexander, K. L. (1988). Factors affecting achievement test scores and marks of black and white first graders. Elementary School Journal, 88 (5), 449-471.

Franklin, R. S., & Resnick, S. (1973). The political economy of racism. New York: Holt, Rinehart & Winston.

Franks, D. J. (1971). Ethnic and social status characteristics of children in EMR and LD classes. Exceptional Children, 37 (7), 537-538.

Fuchs, D. (1987). Effects of examiner familiarity on LD and MR students' language performance. Remedial and Special Education, 8 (4), 47-52.

Fuchs, D., & Fuchs, L. S. (1989a). Effects of examiner familiarity on black, Caucasian, and Hispanic children: A meta-analysis. Exceptional Children, 55, 303-308.

Gamoran, A. (1992). Access to excellence: Assignment to honors English classes in the transition from middle to high school. Education Evaluation and Policy Analysis, 14 (3), 185-204.

Glass, G. V., & Hopkins, K. D. (1996). Statistical methods in education and psychology. Needham Heights, MA: Allyn and Bacon.

Glutting, J. J., Kelly, M. S., Boehm, A. E., & Burnett, T. R. (1989). Stability and predictive validity of the Boehm Test of Basic Concepts-Revised among black kindergartners. Journal of School Psychology, 27, 365-371.

Good, R., Vollmer, M., Creek, R. J., Katz, L., & Chowdhri, S. (1993). Treatment utility of the Kaufman Assessment Battery for Children: Effects of matching instruction and student processing strength. School Psychology Review, 22, 8-26.

Good, T. (1970). Which pupils do teachers call on? Elementary School Journal, 70, 190-198.

Gould, S. J. (1981). The mismeasure of man. New York: W. W. Norton.

Gresham, F. M., & Witt, J. C. (1997). Utility of intelligence tests for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. School Psychology Quarterly, 12 (3), 249-267.

Grossman, F. M., & Franklin, N. K. (1988). Bias effects in speech-language assessment and decision-making. Language, Speech, and Hearing Services in the Schools, 19 (2), 153-159.

Gorsuch, R. L. (1983). Factor analysis (2nd Ed.). Hillsdale, NJ: Erlbaum.

Gutkin, T., & Reynolds, C. (1981). Factoral similarity of the WISC R for white and black children from the standardization sample. Journal of Educational Psychology, 73 (2), 227-231.

Hall, B. (1985). Survey of the technical characteristics of published educational achievement tests. Educational Measurement: Issues and Practices, 4,6-14.

Hargreaves, A. H. (1967). Social relations in secondary school. New York: Humanities Press.

Harry, B. (1992a). Cultural diversity, families, and the special education system: Communication for empowerment. New York: Teachers College Press.

Hartladge, L., & Terzrow, C. (1986). Neuropsychological assessment. Special Services in the Schools, 2 (2-3), 107-119.

Heller, K. A., Holtzman, W. H., & Messick, S. (Eds.) (1982). Placing children in special education: A strategy for equity. Washington, D. C.: National Academy Press.

Herrnstein, R. J., & Murray, C. (1994). The bell curve: Intelligence and class structure in American life. New York: Free Press.

Hilliard, A. (1989). Teachers and cultural styles in a pluralistic society. NEA Today, 7 (6), 65-69.

Hilliard, A. (1991). The learning potential assessment device and instrumental enrichment as a paradigm shift. In A. G. Hilliard, III (Ed.), Testing African American students: A special re-issue of the Negro education reviews (pp. 200-208). Morristown, NJ: Aaron.

Holtzman, W. H. (1985). The influence of sociocultural factors on the achievement and placement of limited English proficient Hispanic students. (ERIC Document Reproduction Serivce No. (ED 261 485)).

Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). Educational and psychological measurement and evaluation (7th Edition). Needham Heights, MA: Allyn & Bacon.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods, 3 (4), 424-453.

Huebner, E. S. (1989). Errors in decision-making: A comparison of school psychologists' interpretations of grade equivalents, percentiles, and deviation IQs. School Psychology Review, 18 (1) 51-55.

Irvine, J. J. (1990). Black students and school failure: Policies, practices, and prescriptions. (ERIC Reproducible Document (ED 320 992)).

Jackson, C. C. (1994). The feasibility of an Afrocentric curriculum. Southern Social Studies Journal, 20 (1), 60-74.

Jenkins, J. R., & Pany, D. (1978). Standardized achievement tests: How useful for special education? Exceptional Children, 44, 448-453.

Jensen, A. (1976). Test bias and construct validity. Phi Delta Kappan, 58 (4), 340-346.

Jensen, A. (1994). Race and sex differences in head size and IQ. Intelligence, 18 (3), 309-333.

Jones, R. L. (Ed.). (1988). Psychoeducational assessment of minority group children: A casebook. Berkeley, CA: Cobb & Henry.

Joreskog, K. G., & Sorbom, D. (1989). LISREL 8: A guide to program and applications. Chicago: SPSS.

Juliano, J., Harrad, A., & Carroll, C. (1988). Three year stability of WISC-R factor scores for black and white, female and male children classified as learning disabled. Journal of School Psychology, 26 (4), 317-325.

Karier, C. J. (1972). Testing for order and control in the corporate state. Educational Theory, 22, 159-180.

Kauffman, J., Wong, K., Lloyd, J., Hung, L., Pullen, P. (1991). What puts pupils at risk? An analysis of classroom teachers' judgments of pupils' behavior. Remedial and Special Education, 12 (5), 7-16.

Kaufman, A. S. (1994). Intelligent testing with the WISC III. New York: Wiley Interscience.

Kaufman, A. S., & Dolppelt, J. E. (1976). Analysis of WISC-R standardization data in terms of stratification variables. Child Development, 47 (1), 165-171.

Kavale, K. (1990). The effectiveness of special education. In T. Gutkin & C. Reynolds (Eds.), The handbook of school psychology (2nd ed., pp. 868-898). New York: Wiley.

Kavale, K., & Forness, S. (1987). How not to specify learning disability: A rejoinder to Koss. Remedial and Special Education, 8, 60-62.

Keith, T. Z., Fugate, M. H., DeGraff, M., Diamond, C. M., Shadrach, E. A., & Stevens, M. L. (1995). Using multi-sample confirmatory factor analysis to test for construct bias: An example using the K-ABC. Journal of Psychoeducational Assessment 13, 347-364.

Keith, T. Z. & Reynolds, C. R. (1990). Measurement and design issues in child assessment research. In C. R. Reynolds & R. W. Kamphaus (Eds.) Handbook of psychological and educational assessment of children: Vol. 1 Intelligence and achievement (pp. 29-61). New York: Guilford Press.

Keith, T. Z. & Witta, E. L. (1997). Hierarchical and cross age confirmatory factor analysis of the WISC III: What does it measure? School Psychology Quarterly, 12 (2), 89-107.

Kelly, D. H. (1972). Track position, school misconduct, and youth deviance. Urban Education, 6, 320-332.

Kershaw, T. (1992). The effects of educational tracking on the social mobility of African Americans. Journal of Black Studies, 23 (1). 152-169.

Kozol, J. (1991). Savage inequalities. New York: HarperCollins.

Kunen, J. (1996, April 29). The end of integration. Time, 39-45.

Kush, J. C., & Watkins, M. W. (1997). Construct validity of the WISC III verbal and performance factors for black special education students. Assessment, 4 (3), 297-304.

Larry P. v. Riles, 343 F. Supp. 1306 (N. D. Cal. 1972) (preliminary injunction). aff'd 502 F. 2d 963 (9th cir. 1974); 495. Supp. 926 (N.D. Cal. 1979) (decision on merits) aff'd (9th cir. no. 80-427, Jan. 23, 1984). Order modifying judgment, C-71-2270 RFP, September 25, 1986.

Lehr, J. B., & Harris, H. W. (1988). At risk, low-achieving students in the classroom: Analysis and action series. (ERIC Reproducible Document (ED 298 232)).

Levin, M. (1994). Comment on the Minnesota transracial adoption survey. Intelligence, 19 (1), 13-44.

Lewontin, R. C. (1992). Biology as ideology: The doctrine of DNA. New York: Harper Perennial.

Link, C. R., & Mulligan, C. R. (1986). The merits of a longer school day. Economic of Education Review, 5 (4), 373-381.

Link, C. R., & Mulligan, C. R. (1991). Classmates' effects on black student achievement in public school classrooms. Economics of Education Review, 10 (4), 297-310.

Look beyond the buses. (1994, May 9). The Saint Petersburg Times, p.1B.

Luge, S., & Luge, R. D. (1978). Relations among teacher rankings, pupil teacher interactions, and academic achievement: A test of the teacher expectancy hypothesis. American Education Research Journal, 15 (4), 103-110.

Maller, S. J., & Braden, J. P. (1993). The construct and criterion related validity of the WISC III with deaf adolescents. In B. A. Bracken (Ed). Monograph series advances in psychoeducational assessment: Wechsler Intelligence Scale for Children: Third Edition; Journal of Psychoeducational Assessment. Brandon, VT: Clinical Psychology Publishing.

Maller, S. J., & Ferron, J. (1997). WISC III factor invariance across deaf and standardization samples. Educational and Psychological Measurement, 57 (6), 987-994.

Macmann, G. M., & Barnett, D. W. (1997). A critical appraisal of intelligent testing with the WISC-III: Introduction to the series. School Psychology Quarterly, 12 (3), 193-196.

MacMillian, D. L., & Balow, I. H. (1988). Impact of Larry P. on educational programs and assessment practices in California. Diagnostique, 17 (1), 57-69.

Marshall et. al v. Georgia. U.S. District Court for the Southern District of Georgia. CV482-233, June 28, 1984: Aff'd (11th cir. no. 84-8771, October 29, 1985).

McCurk, F. (1951). Comparison of the performance of negro and white high school seniors on cultural and non-cultural psychological tests questions. Washington, DC: Catholic University of America Press.

McMillan, D. L., & Rechsley, D. J. (1998). Overrepresentation of minority students: The case for greater specificity or reconsideration of the variables examined. Journal of Special Education, 32 (1), 15-24.

Medina, N. & Neill, D. M. (1990). Fallout from the testing explosion (3rd Edition). Cambridge: Fair Test.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd Edition). New York: Macmillan.

Microsoft Corporation (1998). The Declaration of Independence [CD-ROM]. Microsoft Encarta. Seattle, WA: Author.

Munford, P. R. (1978). A comparison of the WISC and WISC-R on black child psychiatric outpatients. Journal of Clinical Psychology, 34 (4), 938-943.

National Association of School Psychologists (1997). School psychology: Blueprint for training. Bethesda, MD: National Association of School Psychologists.

National Center for Educational Statistics. (1992). Condition of education 1992. Washington, D. C.: U. S. Department of Education.

Oakes, J. (1986). Beyond tracking. Educational Horizons, 65 (1), 32-35.

Oakes, J. (1995). Two cities' tracking and within school segregation. Teachers College Record, 96 (4), 681-690.

Oakes, J. & Wells, A. S. (1998). Detracking for high student achievement. Educational Leadership, 55 (6), 38-41.

Office of Civil Rights. (1998). 20th Annual report to Congress. Washington, DC: Office of Civil Rights.

O'Reilly, C., Northcraft, G. B., & Sabers, D. (1989). The confirmatory bias in special education eligibility decisions. School Psychology Review, 18 (1), 130-135.

Orfield, G., Bachmeier, M. D., James, D. R., & Eitle, T. (1997). Deepening segregation in American public schools: A special report from the Harvard project on school desegregation. Equity and Excellence in Education, 30 (2), 5-24.

Patton, J. (1992). Assessment and identification of African American learners with gifts and talents. Exceptional Children, 59, 150-159.

Patton, J. (1992). Assessment and identification of African American learners with gifts and talents. Exceptional Children, 59 (2), 150-159.

Patton, J. (1998). The disproportional representation of African Americans in special education: Looking behind the curtain for understanding and solutions. Journal of Special Education, 32 (1), 25-31.

PASE (Parents in Action on Special Education) v. Joseph P. Hannon. U. S. District Court, Northern District of Illinois, Eastern Division No. 74 (3586), July 1980. Also 506 F. Supp. 831 (N. D. Ill., 1980).

Peoples, C. E. (1995). The influence of race on 3 year old children's performance on the SB-IV. Intelligence, 21 (1), 69-82.

Poteat, G. M., Wuensch, K. L., & Gregg, N. B. (1988). An investigation of differential prediction with the WISC-R. Journal of School Psychology, 26, 58-68.

Ravitz, M. (1963). The role of the school in the urban setting. In H. A. Passow (Ed.) Education in depressed areas (pp. 201-210). New York: Teachers College, Columbia University.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. Psychological Bulletin, 114 (3), 552-566.

Reschly, D. J. (1987). Assessing educational handicaps. In A. Hess & I. Wiener (Eds.) Handbook of forensic psychology (pp. 155-187). NY: Wiley.

Reschly, D. J. (1991). The effects of placement litigation on psychological and education classification. Diagnostique, 17 (1), 6-20.

Reschly, D. J. (1997). Disproportionate minority representation in general and special education: Patterns, issues and alternatives. (ERIC Reproducible Document (ED 415 632)).

Reschly, D. J., Gresham, F. M. (1989). Current neuropsychological diagnosis of learning problems: A leap of faith. In C. R.Reynolds & E. Fletcher-Janzen (Eds.), Handbook of clinical neuropsychology (pp. 503-519). New York: Plenum.

Reschly, D. J., Kicklighter, R., & McKee, P. (1988). Recent placement litigation, Parts I, II, III. School Psychology Review, 17, 9-50.

Reschly, D. J., & Wilson, M. S. (1995). Cognitive processing versus traditional intelligence: Diagnostic utility, intervention implications, and treatment validity. School Psychology Review, 19 (4) 443-458.

Reilly, T. F. (1991). Cultural bias. The albatross of assessing behavior-disordered children and youth. Preventing School Failure, 36 (1), 50-53.

Reynolds, C. R. (1980). Differential construct validity of intelligence as popularly measured: Correlations of age with raw scores on the WISC-R for Blacks, Whites, Males, and Females. Intelligence, 4 (4), 371-378.

Reynolds, C. R. (1982a). Methods for detecting construct and predictive bias. In R. Berk (Ed.), Handbook of methods for detecting test bias (pp. 275-301). Baltimore: Johns Hopkins University Press.

Reynolds, C. R. (1982b). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology (pp. 178-208). New York: John Wiley.

Reynolds, C. R. (1991). Methods for studying bias in psychological and educational tests. Diagnostique, 17 (1), 21-39.

Reynolds, C. R., & Gutkin, T. B. (1980). Factoral similarity of the WISC-R for Anglos and Chicanos referred for psychological services. Journal of School Psychology, 18 (1), 34-39.

Reynolds, C. R., & Jensen, A. R. (1983). WISC-R subscale patterns of abilities of blacks and whites matched on full scale IQ. Journal of Educational Psychology, 75 (2), 207-214.

Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. Structural Equation Modeling, 3 (4), 369-379.

Rist, R. C. (1970). The socialization of the ghetto child into the urban school system. (ERIC Reproducible Document (ED 111 898)).

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. New York: Holt, Rinehart & Winston.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception processes. Journal of Personality and Social Psychology, 32, 880-892.

Rossmiller, R. A. (1987). Achieving equity and effectiveness in schooling. Journal of Education Finance 12 (4), 561-577.

Rowe, D. C., Vazsonyi, A. T., & Flannery, D. J. (1994). No more than skin deep: Ethnic and racial similarity in developmental process. Psychological Review, 101 (3), 396-413.

Russo, C. J., & Talbert-Johnson, C. (1997). The overrepresentation of African American children in special education: The resegregation of educational programming? Education and Urban Society, 29 (2), 136-148.

Sandoval, J. (1979). The WISC R and internal evidence of test bias with minority groups. Journal of Consulting and Clinical Psychology, 47 (5), 919-926.

Schafer, W. E., & Olexa, C. (1971). Tracking and opportunity: The locked out process and beyond. Scranton, PA: Chandler.

Serwatka, T., Deering, S., & Grant, P. (1995). Disproportionate representation of African Americans in emotionally handicapped classes. Journal of Black Studies, 25 (4), 492-506.

Shapiro, A. S., Benjamin, W. J., & Hunt, J. J. (1995). Curriculum and schooling : A practitioner's guide. Palm Springs, CA : ETC Publications.

Shapiro, E. S., & Derr, T. F. (1987). An examination of overlap between reading curricula and standardized achievement tests. Journal of Special Education, 21 (2), 59-67.

Share, D. L., McGee, R., & Silva, P. D. (1989). IQ and reading progress: A test of the capacity notion of IQ. Journal of the American Academy of Child and Adolescent Psychiatry, 28, 97-100.

Share, D. L., McGee, R., & Silva, P. D. (1991). The author's reply. Journal of the American Academy of Child and Adolescent Psychiatry, 30, 697.

Siegel, L. (1989). IQ is irrelevant to the definition of learning disabilities. Journal of Learning Disabilities, 22, 469-478.

Siegel, L. (1992). Dyslexics vs. poor readers: Is there a difference? Journal of Learning Disabilities, 25, 618-629.

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 (6), 1039-1046.

Slate, J. R., Jones, C. H., Graham, L. S., & Bower, 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 (2), 104-107.

Smedley, A. (1993). Race in North America: Origin and evolution of a world view. San Fransisco, CA: Westview Press.

Smith, G. R. (1983). Desegregation and assignment of children to classes for the mildly retarded and learning disabled. Integrated Education, 21, 208-211.

Snyder, M. (1981). Seek and ye shall find: Testing hypotheses about other people. In E. Higgins, C. Herman, & M. Zanna (Eds.), Social cognition: The Ontario symposium on personality and social psychology. 75-91. Hillsdale: Erlbaum.

Soodak, L. C., & Podell, D. M. (1993). Teacher efficacy and student problem as factors in special education referral. Journal of Special Education, 27 (1), 66-81.

Stedman, L. C. (1994a). Incomplete explanations: The case for U. S. performance in international assessments. Educational Researcher, 23(7), 24-32.

Stedman, L. C. (1995). The new mythology about the status of U. S. schools. Educational Leadership, 52 (5), 80-85.

Stevens, J. (1990). Intermediate statistics: A modern approach. Hillsdale, NJ: Lawrence Erlbaum Associates.

Taylor, R., Ziegler, E., & Partenio, I. (1985). Factor structure of the WISC-R across ethnic groups: An investigation of construct validity. Diagnostique, 11, 9-13.

Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. Higgins, C. Herman , & M. Zanna (Eds.), Social cognition: The Ontario symposium on personality and social psychology. 136-152. Hillsdale: Erlbaum.

Teeter, P. A., & Smith, P. L. (1993). WISC III & WJ R: Predictive and discriminate validity for students with severe emotional disturbance. In B. Bracken (Ed.) Monograph series advances in psychoeducational assessment: Wechsler Intelligence Scale for Children: Third Edition; Journal of Psychoeducational Assessment. Brandon, VT: Clinical Psychology Publishing.

Thuriow, M. L., & Ysseldyke, J. E. (1982). Factors considered important in decision making: a preliminary investigation. Diagnostique, 7, 179-188.

Tindall, G. B. (1988). America: A narrative history (2nd Ed.). New York: W. W. Norton.

Valencia, R. R., Rankin, T., & Oakland, R. (1997). WISC-R factor structures among White, Mexican American, and African American children: A research note. Psychology in the Schools, 34 (1), 11-16.

Vance, H. B., Hankins, T. J., & McGee, B. B. (1979). A preliminary study of Black and White differences on the revised Wechsler Intelligence Scale for Children. Journal of Clinical Psychology, 35 (4), 815-819.

Vance, H. B., & Sabatino, R. (1991). Identifying sources of bias in the WISC R. Diagnostique, 17 (1), 40-48.

Vance, H. B., & Engin, A. (1978). Analysis of cognitive abilities of black children's performance on the WISC R. Journal of Clincial Psychology, 34 (2), 452-456.

Villegas, C., & Watts, T. (1991). Culturally responsive pedagogy for the 1990s and beyond. Trends and Issues Paper No. 6. (ERIC Reproducible Document (ED 339 698)).

Walters, R. (1995). The impact of "Bell Curve" ideology on African American public policy. American Behavioral Scientist, 39 (1), 98-108.

Wagner, T. (1995). What's School Really for, Anyway? And Who Should Decide? Phi Delta Kappan, 76 (5), 393-398.

Ward, S. B., Ward, T. J., & Clark, H. T. (1990). Classification congruence among school psychologist and its relationship to type of referral question and professional experience. Journal of School Psychology, 29, 89-108.

Wechsler, D. (1991). Manual for the Wechsler Intelligence Scale for Children (3rd Edition). San Antonio, TX: The Psychological Corporation.

Webster, S. F. (1974). Education of black Americans. New York: Intext Education Publishers.

Wessel, J., & Potter, A. (1994). Analysis of WISC III data from an urban population of referred children. Paper presented at the annual meeting of the National Association of School Psychologist, Seattle, WA. (ERIC Document Reproduction No. ED 371051).

Wiess, L. G., Prifitera, A., & Roid, G. (1993). The WISC III and the fairness of predicting achievement across ethnic and gender groups. In B. A. Bracken (Ed.), Monograph series advances in psychoeducational assessment: Wechsler Intelligence Scale for Children: Third Edition; Journal of Psychoeducational Assessment. Brandon, VT: Clinical Psychology Publishing.

Witt, J. C., Elliott, S. N., Kramer, J. J., & Gresham, F. M. (1994). Assessment of children: Fundamental methods and practices. Madison, WI: Brown & Benchmark.

Witta, E. L., & Keith, T. Z. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC-III: What does it measure? School Psychology Quarterly, 12 (2), 89-107.

Ysseldyke, J., & Algozzine, B. (1981). Diagnostic classification decisions as a function of referral information. Journal of Special Education, 15 (4), 429-435.

Ysseldyke, J., & Algozzine, B. (1983). LD or not LD: That's not the question. Journal of Learning Disabilities, 16 (1), 29-31.

Ysseldyke, J., Algozzine, B., Regan, R., & McGue, M. (1981). The influence of test scores and naturally occurring pupil characteristics on psychoeducational decision making with children. Journal of School Psychology, 19 (2), 167-177.

Ysseldyke, J., & Mirkin, P. (1983). An analysis of current practice in referring students for psychoeducational evaluation: Implications for change. (ERIC Reproducible Document (ED 228 823)).

Appendix I

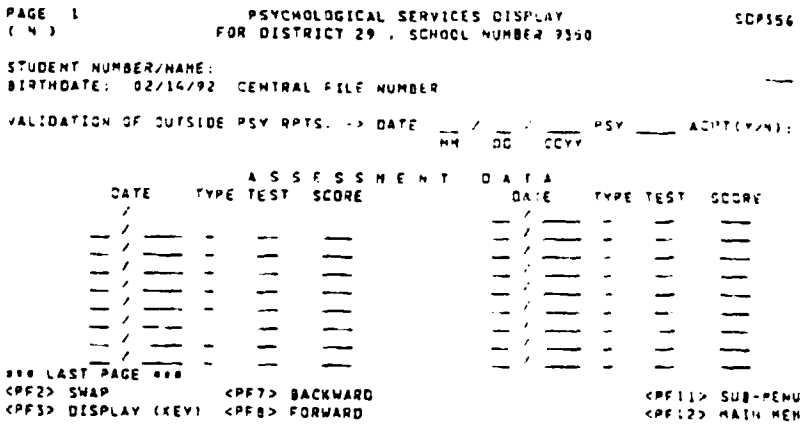
Referral Data Screen

```
PAGE  1                 STUDENT SERVICES REFERRAL DISPLAY                    SDP3563
( L )                  FOR DISTRICT 29 , SCHOOL NUMBER 9350
STUDENT NUMBER/NAME:
BIRTHDATE:               CENTRAL FILE NUMBER: ____                            ___

                  R  E  F  E  R  R  A  L     H  I  S  T  O  R  Y
REFERRAL     REFERRAL    REFERRAL     REFERRAL       SERV  COMPLETION       COMP  WKR
CODE/STAT      DATE    REASON/TYP   SITE/ETH AA    CODE     DATE           SITE  NUM
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
____  _  __ / __ / ____  _ _   _   ____  _  __    __   __ / __ / ____  ____  ____
*** LAST PAGE ***
<PF3> DISPLAY (KEY)   <PF7> BACKWARD                              <PF11> SUB-MENU
                      <PF8> FORWARD                               <PF12> MAIN MENU
```

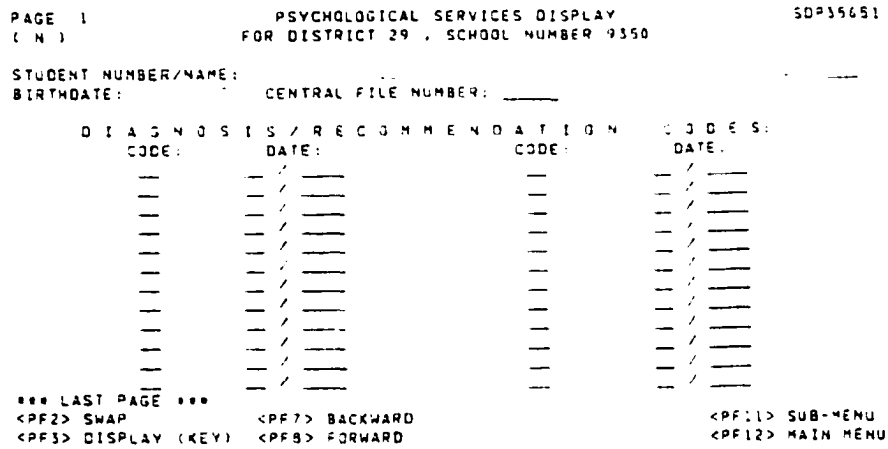
Appendix II

Assessment Data Screen



Appendix III

Recommendations Data Screen



Appendix IV

Reliability Coefficients of the Subtests, IQ Scales, and Factor Based Scales by Age

Table 5.1 Reliability Coefficients of the Subtests, IQ Scales, and Factor-Based Scales, by Age

Subtest/Scale	Age in Years											Average
	6	7	8	9	10	11	12	13	14	15	16	[illegible]
Information	73	76	86	81	82	85	85	85	87	88	88	84
Similarities	81	77	84	80	82	82	84	74	84	81	84	81
Arithmetic	81	73	78	71	79	79	74	81	77	81	82	78
Vocabulary	92	79	88	82	88	88	89	89	91	91	89	87
Comprehension	75	72	85	74	79	78	81	73	76	80	73	77
Digit Span [1]	79	81	84	82	84	84	87	87	84	91	89	85
Picture Completion	78	84	81	80	74	76	72	73	72	82	75	77
Coding	75	70	—[2]	—[2]	78	82	—	—	70	80	—[2]	78
Picture Arrangement	82	84	72	72	74	73	79	76	78	73	73	75
Block Design	82	77	83	85	89	84	87	90	90	92	90	87
Object Assembly	71	65	65	75	69	65	68	75	80	76	71	69
Symbol Search	58	76	—	—[2]	72	79	—[2]	—[2]	75	82	—[2]	75
Mazes	80	78	75	66	70	68	66	73	70	61	67	70
Verbal	93	92	96	93	95	95	95	94	95	96	95	95
Performance	91	90	90	91	91	90	91	90	89	94	92	91
Full Scale	95	94	96	95	96	95	96	95	95	97	96	96
Verbal Comprehension	91	91	95	93	94	94	95	93	95	95	94	94
Perceptual Organization	91	90	89	90	90	89	91	91	89	93	90	90
Freedom-Distractibility	87	84	87	83	86	88	86	88	86	91	90	87
Processing Speed	81	80	84	85	84	87	87	82	82	91	91	85

[illegible]

[1] The median [illegible]

[2] For Coding and Symbol Search, [illegible]

Appendix V

Sample Research Datasheet

Central File Number Ethnicity Gender Age WISC III Full Scale Score VIQ PIQ PC Inf Cod Sim PA BD Voc OA Comp

About the Author

Dr. Cametra Reed is a graduate of the School Psychology Program at the University of South Florida. Dr. Reed received her Bachelor of Science degree from the University of Central Florida in Orlando, FL. Current interests and experiences emphasize multicultural education issues and children with emotional, behavioral and social challenges. Dr. Reed currently directs the Village of Excellence Academy, a charter school that targets children who have been frequently expelled or suspended from the public school system. Dr. Reed is a native of and resides in the Tampa Bay area.