To the Graduate Council:

I am submitting herewith a dissertation written by LaRonta Michelle Upson entitled "Effects of an increasingly precise socioeconomic match on mean score differences in nonverbal intelligence test scores." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

R. Steve McCallum
Major Professor

We have read this dissertation
and recommend its acceptance:

Thomas George

Charles Hargis

William Calhoun

Bruce A. Bracken

Acceptance for the Council:

Dr. Anne Mayhew
Vice Provost and Dean of Graduate Studies

(Original signatures are on file with official student records.)

EFFECTS OF AN INCREASINGLY PRECISE SOCIOECONOMIC MATCH ON MEAN SCORE DIFFERENCES IN NONVERBAL INTELLIGENCE TEST SCORES

A Dissertation

Presented for the

Doctor of Philosophy Degree

The University of Tennessee, Knoxville

LaRonta Michelle Upson

August 2004

DEDICATION

This dissertation is dedicated to my parents, Laronza Sr. and Gail Upson. Your love and support has allowed me the confidence to accomplish many things in my life. You have taught me by your example that character, a strong work ethic, and faith in God are essential to achieving my greatest goals. I am so grateful to have you as my parents and as my friends. Thank you.

This dissertation is also dedicated to my sister LaTrisha and brother Laronza Upson Jr. You two are my best friends in the world and my greatest inspiration. Through it all, the most difficult part of completing this degree was being away from the two of you. When we are together we talk, laugh, argue, and laugh again just as if we were never apart. I am so proud of the both of you and your accomplishments. Thank you for always supporting me.

ACKNOWLEDGEMENTS

I would like to express my gratitude to several people who have supported my efforts in completing this project. First, I would like to thank my advisor and committee chair, Dr. Steve McCallum for his support and insight throughout my graduate studies and particularly, during the completion of my dissertation. I would also like to thank my committee members, Dr. Thomas George, Dr. Charles Hargis, Dr. William Calhoun, and Dr. Bruce Bracken for their important contributions towards completion of this project.

Additionally, I would like to thank Dr. Christopher Skinner for his support during my time at the University of Tennessee and for helping me to expand my experiences in applied research and consultation in school psychology. I have learned a lot from your experiences. Thank you, Millie Cheatham, for your guidance and friendship when I needed it most. I will truly miss you. I am deeply appreciative of Dr. Lynn Cagle for his support and for selecting me as a Holmes scholar, which allowed me to make life-long friends and work professionally with other minority Ph.D. students from across the country. I would also like to thank Mike O'Neil for his statistical expertise.

Finally, I am deeply indebted to several people for their support throughout this process including my parents Laronza and Gail Upson, my sister and brother, LaTrisha and Laronza Jr., and my colleagues and friends, Dionne Smith (I could not have done this without all your help), Shelton Starks, and Shannon Sharp. Above all, I would like to thank God, for His love, comfort, strength, and faithfulness everyday of my life.

ABSTRACT

Data consisting of 3,861 participants from the Universal Nonverbal Intelligence Test (UNIT) standardization and validity studies samples were used to examine the effects of increased socioeconomic refinement on cultural and racial mean difference scores. Groups were equated not only on age, sex, and highest parent education level, as reported in the UNIT manual, but on two additional socioeconomic status (SES) indicators: community setting and both parent education levels. Results suggest that additional refinement on socioeconomic variables does little to further reduce mean score IQ differences in UNIT Standard FSIQ scores between African Americans and Whites (n=168 in each group; mean difference = 8.51, effect size= .55); however, the 8.51 mean difference is considerably smaller than the 15-point difference typically observed between African American and Whites and is lower than the 11 point difference shown for WISC-III FSIQ scores, even after SES matching. There were no significant mean IQ score differences (n=162, p>.05) between Hispanics and Non Hispanics indicating that additional socioeconomic status refinement does contribute to reductions in mean score differences in IQ between these populations. In fact, Hispanics scored higher than their Non Hispanic counterparts on several subscales. Findings offer support for the use of the UNIT with diverse populations, as this measure of intelligence may limit the influence of irrelevant cultural factors in assessment. Future research on the use of nonverbal intelligence measures to predict minority student achievement, progressive conceptualizations of intelligent behavior, and exploration of within racial-ethnic group factors that contribute to or inhibit cognitive growth and academic achievement in minorities is warranted.

TABLE OF CONTENTS

1. INTRODUCTION…….…….. 1

Purpose……. 1

Brief Historical Perspectives on Intelligence……... 1

Factors Influencing Performance on Intelligence Tests…… 5

Potential Sources of Bias…….. 13

Fairness in Testing……..... 17

Mean Score Differences as a Fairness Issue..……... 20

Statement of the Problem…….. 22

Research Questions……... 23

2. METHODS…….. 24

Participants and Procedures…….. 24

Instrument……. 25

Data Analysis…… 27

3. RESULTS…… 29

Research Question One……. 30

Research Question Two……. 31

Research Question Three..……. 32

Research Question Four.…… 33

Research Question Five.……. 34

Research Question Six…..……. 35

4. DISCUSSION…….. 36

Mean Score Differences in Demographically Matched Samples…….. 37

Socioeconomic Status as a Predictor of Intellectual Ability……. 40

Comparative Study and Future Directions…… 41

Limitations and Implications for Future Research……. 43

REFERENCES…….. 48

APPENDICES…… 60

Appendix A: Tables…….. 61

Appendix B: Copy of Accepted IRB Application…… 67

Appendix C: Copy of Permission from Riverside Publishing Co……. 71

VITA…….. 74

LIST OF TABLES

Table Page

1. UNIT performance by African Americans and a demographically matched White sample…….. 62
2. UNIT performance by Hispanics and a demographically matched Non Hispanic sample…….. 63
3. Summary of African American and White FSIQ scores on Abbreviated, Standard, and Extended Batteries when matched on age, sex, highest parent education level (Manual); age, sex, community setting, highest parent education level; and age, sex, community setting, and both parents education levels…… 64
4. Summary of Hispanic and Non Hispanic FSIQ scores on Abbreviated, Standard, and Extended Batteries when matched on age, sex, highest parent education level (Manual); age, sex, community setting, highest parent education level; and age, sex, community setting, and both parents education levels……. 65
5. Means, standard deviations, and difference FSIQ scores based on follow-up matching procedures for African Americans and Whites… 66
6. Means, standard deviations, and difference FSIQ scores based on follow-up matching procedures for Hispanics and Non Hispanics ….. 66

CHAPTER 1

Introduction

Purpose

Administration of the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998) results in relatively small minority-majority population mean score IQ differences when participants are matched on age, sex, and highest parent education attainment--a difference that is smaller than most other cognitive measures. The purpose of this study is to evaluate the effects of further refining the match by equating groups not only on age, sex, and highest parent education level, but two additional socioeconomic indicators: community setting and both parent education levels. Will this additional refinement further reduce the mean score differences that exists between groups (e.g., Whites vs. African Americans and Non Hispanics vs. Hispanics)?

Brief Historical Perspectives on Intelligence

A number of important and sometimes contradictory theories explain the form and nature of intelligence. Some theorists tend to emphasize a dominant, general intelligence construct (Jensen, 1981; Spearman, 1927) said to undergird all mental activity, whereas others conceptualize intelligence as manifesting itself in a number of specific and largely independent domains (Gardner, 1983; Sternberg, 1985; Thurstone, 1960). According to Spearman (1927), intelligence is composed of both a dominant general factor (g) and a number of specific factors. Spearman's *'g'* is used to explain individual differences in performance on intelligence tests. Presumably, all tests of mental ability measure to some degree this *'g'* factor, and any difference in performance that exists between individuals and groups are attributable primarily to differences in *'g'*.

The Cattell-Horn (Cattell, 1971) theory provides a much more recent model that retains some of the characteristics of Spearman's conceptualization. According to this model, mental ability can be further disaggregated to reflect two broad subtypes, namely "fluid" and "crystallized" intelligence. Fluid intelligence reflects one's information processing ability and the relatively unlearned ability to solve novel problems. Crystallized intelligence refers to those abilities that are dependent on acquired knowledge, and are more influenced by experience and education. This theory has been revised extensively and is now referred to as the Cattell-Horn-Carroll (CHC) model, integrating both the original Cattell-Horn Gf-Gc theory and Carroll's three-stratum model of cognitive abilities (see Horn & Noll, 1997; Carroll, 1993). Horn (1991) expanded upon the original two factor Gf-Gc model to include nine broad cognitive abilities: Fluid Intelligence (Gf), Crystallized Intelligence (Gc), Short-Term Acquisition and Retrieval (Gsm), Visual Intelligence (Gv), Auditory Intelligence (Ga), Long-Term Storage and Retrieval (Glr), Cognitive Processing Speed (Gs), Correct Decision Speed (CDS), and Quantitative Knowledge (Gq). Carroll proposed a three-stratum model to include approximately 70 additional subconstructs that fall under eight broad cognitive ability factors-- factors very similar to those proposed by Horn. These include Fluid Intelligence (Gf), Crystallized Intelligence (Gc), General Memory and Learning (Gy), Broad Visual Perception (Gv), Broad Auditory Perception (Gu), Broad Retrieval Ability (Gr), Broad Cognitive Speediness (Gs), and Processing Speed/Reaction Time Decision Speed (Gt).

This empirically based model is derived from factor analysis of large data sets obtained from the administration of various cognitive tasks. Moreover, a number of intelligence tests have been developed using the Cattell-Horn model as their theoretical

basis (Flanagan & Ortiz, 2002). An underlying principle of the Kaufman Assessment Battery for Children (KAB-C; Kaufman & Kaufman, 1983), for example, is the Cattell-Horn theory of fluid and crystallized intelligence. Because intelligence tests are essentially a measure of what one has learned (Kaufman, 1990), the crystallized scores are particularly susceptible to the adverse influence of poor achievement skills, limited vocabulary, and limited general knowledge (Kaufman, 1994), as often seen in low-income minority students.

Other multi-component theories of intelligence are based on cognitive, information processing, or biopsychological models. These theories typically posit the idea of more than two "intelligences" as suggested by Spearman (1927), but they disagree on the nature and number of these intelligences. For example, Guilford's (1967) "Structure of Intellect" theory initially proposed 120 intellectual abilities which are categorized these along three independent dimensions: activities or operations (i.e., cognition, memory), content on which operations are performed (i.e., visual, auditory, symbolic), and the product or result of applying particular operations (i.e., units, classes, implications). On the other hand, Gardner (1983) originally suggested the existence of seven relatively independent intelligences including linguistic, musical, logical-mathematical, spatial, body-kinesthetic, interpersonal, and intrapersonal, and later added "natural" intelligence (as cited in McDevitt & Ormrod, 2002). He contends that psychometric batteries tap only linguistic, logical, and to some degree spatial intelligences but don't account for other forms of intelligence (Neisser et al., 1996). Although he offers little empirical and psychometric support for his ideas, Gardner's

theory remains influential in understanding the processes that underlie intelligence, at least from a multiple intelligence perspective.

Like Gardner, Sternberg's (1985) triarchic theory of intelligence conceptualizes intelligence beyond what is typically measured by intelligence tests. Sternberg identifies three subtheories that help explain his triarchic model of intelligence: contextual theory, experiential theory, and componential theory. The contextual subtheory describes behaviors that demonstrate intellectual competence in different environments. The experiential subtheory refers to intelligence that requires the application of learned knowledge and skills to new situations. Finally, the componential theory involves cognitive processes, specifically information processing and problem solving. This theory suggests that there are three interrelated aspects of intelligence: analytic (book smarts), creative, and practical (street smarts). Of Sternberg's three intelligences, only analytical ability is assessed by conventional intelligence tests (Neisser et al., 1996). However, Sternberg and Clinkenbeard (1995) suggests that diverse students manifesting practical intelligence can perform in ways that are just as predictive of grades as conventional tests if they are taught in a way that capitalizes on their practical strengths. As such, these equally important forms of ability (i.e., practical and creative ability), as often demonstrated in ethnically and socioeconomically diverse students, may go undiscovered by conventional intelligence measures. That is, members of minority groups may perform as well as or better than their majority group counterparts on these indices (Torrance, 1971).

Factors Influencing Performance on Intelligence Tests

For decades, the assessment of intelligence has been the subject of much research and controversial debate. In 1905 Alfred Binet published the first practical "intelligence test," presumably to identify low-performing students who could not succeed in regular classes and who might need some form of specialized services (Hilliard, 1989; Jensen, 1981). Since that time, "IQ tests," referring to the quotient derived from them (Hilliard, 1989), have been used to provide predictive information about educational and vocational success (Copeland, Conrad, Chansky, 1978; Wagner, 1997).

Theoretically, measures of intellectual ability provide predictive information about future academic achievement, can be used to diagnosis learning difficulties and mental deficits, and when necessary can help to identify the need for specialized services outside the scope of that provided by general education programs. Various editions of the Wechsler Intelligence Scales and the Stanford Binet have traditionally been the more preferred tests (Jensen, 1981) and thus, the most extensively studied. Even now, the current versions of these tests remain the most commonly used for assessing the intellectual ability of students suspected of learning difficulty, mental retardation, and giftedness. However, the use of the Wechsler and Binet scales in the schools has raised a number of controversial issues related to their appropriateness for use with diverse populations, in part because of the relatively large mean IQ score differences among those groups. Decades of research, in both educational and psychological literature, have helped to explain factors that influence IQ scores including culture, language, schooling and education, socio-economic status, race and ethnicity, and genetics. The fact that some

groups have experiences that are markedly different from mainstream culture has been one of the most frequently cited explanations to date (e.g., Helms, 1992).

Culture and language

Much of the controversy raised concerning testing diverse populations focuses on the issue of bias and fairness, with many researchers questioning the validity of scores traditional intelligence tests yield with culturally and linguistically diverse populations (Helms, 1992; Helms, 2002; Hilliard, 1983). Presumably, traditional intelligence tests rely too heavily on verbal skills and factual knowledge, which may unfairly penalize students with non-dominant cultural and language experiences and result in the interpretation of cultural or language differences as disabilities. As a result, the degree of cultural and linguistic "loading" on intelligence test measures has been cited as contributing to differential test performance among groups (Dana, 1993; Suzuki & Valencia, 1997). Taylor and Lee (1995) assert that differential performance is partly the result of standardized testing rules, which are based on a number of faulty assumptions created in an effort to control or standardize the testing environment. They note that these strict controls placed on the testing environment are often incompatible with the various styles of communication that individuals may bring into the testing situation.

One assumption identified by Taylor and Lee (1995) as inherent in many standardized tests, is the belief "that all individuals communicate their experiences in a similar manner" (p. 41) when in fact differences in expressive language and cognitive problem solving style may vary not only across cultures but with each individual. It is precisely this variation that often interferes with the validity of scores derived from traditional IQ tests. Sturn and Johnston's (1999) investigation of children with specific

language impairments (SLI) found that greater use of private speech in SLI students was associated with less cognitive efficiency, as indicated by IQ scores. Thus, to presuppose that all individuals convey knowledge in the same way, fails to consider that equally intelligent people may use differing, but equally intelligent reasoning strategies (Helms, 1992).

A second assumption of conventional intelligence tests generally accepts the premise that there is one right answer to a given question or problem, and the individual who is able to access these predetermined answers, which are often grounded in common American culture, is therefore "intelligent" (Helms, 1992; Hilliard, 2002). Consequently, many argue that results of standardized tests "favor children who speak common American English, primarily because these students are able to respond to questions that represent a familiar language based on familiar experiences" (Hilliard, 2002, p. 98). Ideally, the use of a variety of acceptable answers would be good practice; however, the very nature of "standardized" test administration often precludes the incorporation of variability in responses (Taylor & Lee, 1995).

Finally, many intelligence tests are based on the fundamental notion of universality, which allows for interpretation and comparison across groups. Hilliard (1983) challenges this assumption of universality by exploring the criteria for item selection on vocabulary subtests for example, which are included in many commonly used intelligence tests. Specifically, he challenges the notion that there is a universal vocabulary that all Americans have had an equal chance of exposure to learn irrespective of culture, home environment, socioeconomic status, or regional differences.

Historically, the idea that tests should not be linguistically or culturally discriminatory has been upheld by several important court decisions and legislative mandates including Larry P. v. Riles (1972), Diana v. State Board of Education (1970), and ultimately the Individuals with Disabilities Education Act (IDEA) or Public Law 105-17 (Taylor & Lee, 1995). Specifically, IDEA states that:

1) tests and other evaluation materials must be selected and administered so as not to be racially or culturally discriminatory.
2) tests and materials must be administered in the child's native language or other mode of communication in which the child is most proficient
3) test materials and procedures must be selected and administered to a child with limited English proficiency or other form of communication disability, so as to reflect accurately the child's ability rather than measuring the child's English language skills or impaired communication skill (Taylor & Lee, 1995).

Despite the requirements of federal law and the aforementioned court decisions, most standardized tests still include culturally-based communication requirements that may confound the results of psychological testing.

Schooling and education

Schools have the potential to impact intelligence in a number of ways. Many of the skills required to perform well on conventional intelligence tests (i.e., factual knowledge, vocabulary, abstract thinking) are learned in school. Thus, scores obtained on IQ tests may vary as a function of educational opportunity (Ceci & Williams, 1997). IQ scores are positively enhanced by schooling experiences that consistently promote

intellectual development and practice in skills similar to those tested on intelligence measures. It is precisely this similarity with school curriculum that IQ tests are able to predict achievement as well as they do (Neisser et al., 1996).

A second line of research suggests that the racial and socioeconomic make-up of schools and communities influence the educational experiences of minority children (Brooks-Gunn, Duncan, Klebanov, & Sealand, 1993). Consequently, some researchers have studied the composition of schools and the problem of disparate educational opportunities (Schofield, 1991; Wells & Crain, 1994). Tracking practices, for example, often relegate minority and low income students to instructional classes where they receive fewer resources and less qualified teachers whereas majority, middle class students are advantaged by more resources, more qualified teachers, and a college preparatory curriculum (Oakes, Gamoran, & Page, 1992). Harnqvist (1968) conducted one of the earliest studies examining the effects of track assignment on IQ. He found students assigned to academic tracks had higher IQ test scores, by approximately 0.62 standard deviations, than those assigned to vocational tracks. These inequitable learning opportunities potentially dictate which students receive quality education and those that do not (Darling-Hammond, 1985) perhaps perpetuating the achievement gaps among majority and minority groups.

Socio-economic status (SES)

SES has been found to consistently predict intellectual performance (Oakland, 1978; Suzuki & Valencia, 1997). One of the earliest studies documenting SES differences in intellectual performance was conducted by Arlitt (1921), who ultimately concluded

that SES transcended race in predicting intelligence test scores. Arlitt found that groups similar on high levels of SES also scored higher on intelligence tests, irrespective of race.

Wechsler (1971) concedes that lower socio-economic groups score lower on IQ tests and suggests that any solution to remedy these group differences ameliorate the social and economic conditions of members within these groups rather than the test itself. Students from low SES backgrounds experience a myriad of environmental, health, and family related problems associated with poverty that may further depress their intelligence. Inadequate nutrition, poor pre- and postnatal care, and limited parental education (McDevitt & Ormrod, 2002), are a few factors that appear to impede the intellectual development of economically disadvantaged students. Matching groups on SES has been shown to reduce majority-minority differences (Prifitera, Weiss, and Saklofske, 1998), and will be a major focus of this study, as discussed later.

Race and ethnicity

Although often used interchangeably, Wijeyesinghe, Griffen, & Love (1997) distinguish between race and ethnicity. Race is defined as "a social construct that artificially divides people into distinct groups based on characteristics such as physical appearance (particularly color), ancestral heritage, cultural affiliation, cultural history, ethnic classification and the social, economic, and political needs of a society at a given period of time" (p. 88). Ethnicity, also a social construct, is defined as "smaller social groups based on characteristics such as a shared sense of group membership, values, behavioral patterns, language, political and economic interests, history and ancestral geographical base" (p. 88). It is difficult to separate the issue of racial-ethnic group differences in IQ test scores from SES differences since minority groups are so

disproportionately represented in the lower socio-economic stratum (Neisser et al., 1996). Thus, any potential influence of race and ethnicity may be inextricably tied to differences in SES as well. Although the early debates focused on 'race' differences in test performance, more progressive ideas suggest that it is actually cultural influence that has the greater impact. Put simply, the cultural experience associated with the 'race' rather than the race itself probably explains the variability in scores (Puente & Salazar, 1998).

On average, racial and ethnic minority groups differ markedly in their performance on intelligence tests. African Americans, on average score approximately 15-points lower (one standard deviation) below that of Whites with the Hispanic-White difference being less pronounced (Neisser et al., 1996). One exception is the Asian population, who on average, score one to five points higher than Whites (Flynn, 1991). The origin of these racial and ethnic group differences have been debated both empirically and emotionally for decades. Researchers have explained these differences using genetic (Herrnstein & Murray, 1994), environmental, socioeconomic (Brooks-Gunn et al., 1993), and historical discriminatory practice (Ogbu, 1994) arguments.

Genetics and environment

The relative contribution of genes and environment on individual differences in intelligence is difficult to determine. Much of the evidence provided in support of a genetic influence has been conducted using twin studies as well as adoption studies. For example, identical twins reared apart have more similar IQs than fraternal twins raised within the same home (Bouchard & McGue, 1981). Additionally, studies show that adopted children have IQs more similar to their biological than their adoptive parents, suggesting a stronger genetic influence. Perhaps the most controversial genetic evidence

offered to explain group differences in intellectual performance was provided in Herrnstein and Murray's (1994) *The Bell Curve*. This text suggests that 60% of the variance observed in test scores is inherited. In general, research has not supported the assertion that mean score differences in IQ are attributable to genetic factors to the extent suggested by Hernstein and Murray.

Scholars have used similar types of evidence (provided to support the idea of genetic determinants of intelligence) to argue a stronger environmental influence. Studies comparing twins reared in the same versus different homes have found that those reared in different homes tend to have less similar IQs than twins raised together (McDevitt & Ormrod, 2002). In a classic study by Scarr and Weinberg (1976), children born of poor biological parents were adopted by middle-class parents with above average IQs and then compared with a similar group of students who remained with their biological parent. Individuals raised by adoptive parents possessed IQs approximately 15 points higher than that of the non adopted group. Most researchers agree that both genetic and environmental influences impact intelligence, but may disagree on the proportional contribution of each.

As previously mentioned, some experts argue that group differences result from test bias (i.e., tests are biased against the lower scoring group). Clearly, systematic differences as a function of group membership, whether due to SES, cultural, or language differences, raises concerns and propels scholars to investigate and explain those differences. However, mean differences do not necessarily indicate the presence of bias. For example, Jensen (1980) notes that there is no a priori reason to believe that all groups should score the same on IQ tests, and to believe that groups should be equal leads to the

"egalitarian fallacy." This is the idea that if a test yields mean score differences between groups, it is therefore biased. Many etiological perspectives and definitions on bias have been published and then disputed in the psychological assessment literature. The following definitions are the most common.

Potential Sources of Bias

The ongoing controversy concerning bias in psychological tests has paralleled that of the classic nature/nurture debate. The term 'bias' has been defined in the literature as systematic error in the measurement of a psychological construct as a function of membership in a cultural or racial subgroup (Reynolds, 1982). Therefore, the idea of "bias" has been conceptualized by many experts as resulting from the use of psychometrically flawed intelligence tests which artificially give the appearance of group differences. Cultural test bias, a term used frequently throughout the test bias literature, historically referred to the idea that tests yield different mean scores across racial groups, resulting in disparate identification and placement of those groups (Helms, 1992). This perspective maintains that some "construct irrelevant" factor, such as language or culture, confound the assessment process and subsequent interpretation of test results.

Inappropriate standardization sample

Bias arguments have been based on the idea that tests are more appropriately used with groups adequately represented in the norming sample. Harrington (1975) suggested that the greater the minority representation in the standardization sample the greater the overall mean score for that particular group representing the greater minority population. He further argued that tests normed primarily on the majority population (i.e., Whites) would not have the same predictive validity for minority groups given their small

representation in the standardization sample. Twenty years later, Fan, Willson, and Kapes (1996) examined Harrington's original conjecture under two distinct construction models: one with differential representation of ethnic groups, the other with maximal representation of one ethnic group. Their results indicated no systematic advantage or disadvantage in test performance as a function of ethnic group representation in the sample; even with 100% ethnic group representation in the sample. Consequently, their results fail to support bias arguments based on the idea that tests favor those most widely represented in the test standardization sample.

Content validity bias

The earliest work in cultural test bias was centered primarily on the issue of content validity bias (Reynolds, Lowe, & Saenz, 1999). For decades, researchers have studied this phenomenon statistically (i.e., Camilli & Shepard, 1987; Nandakumar, Glutting, & Oakland, 1993) and by more subjective analyses, specifically visual inspection of potentially biased items by members of minority groups (Jensen, 1976; Kaufman, 1979). Reynolds et al. (1999) state that a test item is considered to be biased in content if (a) the item requires knowledge or skills that the examinee has not had the opportunity to learn; (b) the language of the item is delivered in such a way that confuses the ethnic minority and an inaccurate understanding of the question may result in erroneous answers; and finally, (c) item scoring, specifically, what is considered correct and incorrect, is based on a sample of primarily majority, American culture, which may unfairly penalize ethnic minorities for an answer that might be correct from their own cultural frame of reference. Some researchers contend that the inclusion of items meeting one or more of the above criteria will constitute content validity bias for ethnic minority

groups, testing their familiarity with American culture rather than the intended construct of intelligence. Consequently, those items found objectionable should be eliminated. However, some say that discarding potentially biased items will do little, if anything, to reduce the mean score differences that exists between groups (Flaugher, 1978; Jensen, 1976). After an extensive review of the literature, Reynolds et al. (1999) conclude that the evidence suggests no systematic bias against minorities due to test content; and when instances of bias do occur they account for less than 5% of the variance in test scores.

Construct validity bias

Construct validity refers to the degree to which a test in fact measures the trait it purports to measure. Thus, a test measuring a different construct as a function of group membership may be considered invalid. One approach used to assess construct validity involves the statistical technique of factor analysis. Factor analysis allows the experimenter to determine the extent to which the same constructs are measured across populations. Hilliard (1979) has offered several perspectives on bias and notes that:

> If the IQ test is a valid and reliable test of 'innate' ability or abilities, then the factors which emerge on a given test should be the same from one population to another, since 'intelligence' is asserted to be a set of mental processes (p. 53).

Reynolds et al. (1999) discusses in detail several studies comparing the factor structure of the Weschler scales across races. Their conclusion supports the similarity of the factor structure across race; they reject the claim of construct validity bias.

Predictive validity bias

A substantial amount of research has been conducted on the differential predictive validity of intelligence tests across race. Predictive validity, said to be the most critical of

all forms of validity in the test bias literature (Reynolds et al., 1999), refers to the relationship between the test score, in this case, and some outcome of future performance (Sattler, 2001). Predictive validity is especially important when results have high interpretive significance and are used to assign students into special education classes, decide admittance into a college or university, or determine competency to perform certain jobs. Bias exists with regard to predictive validity when a test score consistently leads to erroneous inferences or predictions as a function of group membership (Cleary, Humphreys, Kendrick, & Wesman, 1975).

Sattler (1974) and more recently Reynolds et al. (1999) have reviewed a number of studies evaluating the differential predictive validity of IQ tests across race and gender. The preponderance of their evidence suggests equivalent validity of IQ measures across groups. Reynold's (1980) found that, in such cases where bias did exist, it acted in favor of Blacks, tending to over-predict their performance. Conclusions reached in these "bias" studies have not been accepted by all researchers who are still convinced of cultural bias in psychological tests (Helms, 1992; Scheuneman, 1987). Helms (1992), for example, argued that Blacks and other minorities are inherently different, cognitively and culturally, than Whites. She offers a number of suggestions to address this cultural issue including the need to determine the cultural equivalency of cognitive ability tests and development of separate racial group norms for existing tests, to name a few.

Disproportionate placement as "bias"

Some experts maintain that the use of "biased" intelligence tests consistently lead to the over-identification of cultural and linguistic minorities (who earn lower IQs) in special education. For over 30 years, the problem of disproportionate representation of

minority students in special education has been a controversial issue (Dunn, 1968; Kaufman, Hallahan, & Ford, 1998) due in part to the inappropriate use and interpretation of standardized tests (Reschly, 1981). Graham and Harris (1989) suggest that the problem of disproportionate representation is not due solely to the inappropriate use of IQ tests but rather socio-political determinants such as lack of flexibility in funding resources, disparate methods in teacher training, in combination with a host of other complexities that in the aggregate add to the problem of overrepresentation. Although, IQ tests do not solely determine eligibility and placement decisions, they do figure significantly in decision- making (Jensen, 1981). Insofar as this problem of over-identification reflects systematic bias in the educational system, rather than authentically higher rates of disability for some groups, efforts should be made to improve the procedure by which students are identified for special education (Oswald, Coutinho, & Best, 2000).

Fairness in Testing

Recently, there has been a decline in test bias research likely due to the development of sophisticated methods for minimizing bias in most well developed tests (Suzuki & Valencia, 1997). Instead, the test bias debate has evolved into discussions of enhanced fairness. In fact, the issue of fairness represents the positive "flip-side" of bias discussions. A "fair" intelligence test is assumed to be one that includes state of the art mechanisms to reduce bias and one that minimizes the influence of construct irrelevant factors (McCallum, 1999). Many nonverbal measures of intelligence, including the UNIT, were developed for the purpose of maximizing fairness in testing for all individuals irrespective of age, sex, race, ethnicity, or language. Typically, nonverbal tests assess an individual's ability to solve problems using memory and reasoning with

visual- spatial tasks, matrices, and pictures that represent universally familiar objects or events (McCallum, 1999). Nonverbal intelligence tests are less dependent upon expressive and receptive language skills and therefore, are especially useful with cultural and language minorities and individuals with limited English proficiency. The implication suggests that intelligence tests, which are able to avoid construct irrelevant language influences by eliminating this requirement altogether, are fairer for those whom language related expectations place them at a disadvantage. These and other important criteria of fairness are discussed below.

McCallum (1999) has identified a number of criteria for reducing test bias and ensuring fairness in assessment. First, and perhaps most importantly, he argues that a language-free test is less susceptible to the language related biases inherent in many traditional language-loaded tests. Second, the unidimensional nature of many existing nonverbal intelligence tests, limit their effectiveness in adequately assessing broad cognitive functioning; a limitation, which initially prompted the development of the UNIT. Multifaceted intelligence tests that require complex reasoning and problem solving skills along a number of dimensions are considered to be a better measure of "g" and thus, fairer (Bracken & McCallum, 1998).

As stated earlier, test items may be considered biased in content if the item requires knowledge or skills that the examinee has not had the opportunity to learn. As a result, poor educational experiences may result in lower levels of acquired knowledge. A fair intelligence test is one that is less focused on crystallized forms of intelligence, which are more sensitive to poor schooling experiences and limited learning opportunities.

Conventional intelligence tests, the verbal scales in particular, have been criticized for relying too heavily on acquired knowledge in the assessment of intelligence.

The use of timed tasks in assessment has been criticized for placing too much emphasis on speed rather than accuracy in the measurement of intelligence (Bracken & McCallum, 1998). Specifically, performance on speeded tasks may place students with processing speed deficits or motor impairments at a disadvantage, if this performance figures substantially into the overall IQ score. McCallum (1999) suggests that tests, which minimize the influence speed, are fairer than those that place great emphasis on this skill.

Reducing bias and enhancing fairness must also include mechanisms for counteracting validity related bias issues. The development of the UNIT, for example, used item bias analyses as well as expert judges to eliminate potential sources of content related bias. Moreover, factor analytic techniques and correlations were established to verify the UNIT as a fair measurement of the intended construct of intelligence across populations. Differential predictive validity of minority groups as opposed to majority groups has been consistently implicated as a source of bias in testing. Statistical analyses were used during the development of the UNIT to ensure the predictability of the UNIT similarly across populations (McCallum, 1999).

Sattler (2001) maintains that all tests are culturally biased to some degree. The extent to which a test can be considered “fair” begins early in the test development phase and rests largely with the tests internal and external psychometric properties and the developer’s foresight to carefully plan for and curtail potential construct irrelevant variables that may impact test performance. Among the many indicators of fairness,

reduction of mean score differences among groups, is the most salient. Tests that show reduced mean score differences between minority and majority groups are assumed to be less biased against the lower scoring group and thus more culturally fair.

Mean Score Differences as a Fairness Issue

Many experts agree that most recently developed tests meet few if any of the psychometric criteria indicative of bias and are more "fair" than previous measures. However, the controversy continues, in part because of the significant mean score differences observed between minority and majority groups. Eventhough, the "mean score difference as bias" argument has been criticized and rejected in psychological assessment research (Jensen, 1981; Reynolds et al., 1999), it is still considered to be a "red flag." When it does occur, it raises doubt among examiners who must administer and interpret test results with confidence. Discussions of bias and fairness would not be relevant if mean differences among groups did not exist. Examining the nature of mean score differences, and reasons for those differences, will help address the larger issue, which is fair, accurate, and valid intellectual assessment of cultural and language minorities.

One approach typically used to examine (and reduce) mean IQ score differences among groups has been to match the two populations in question on a number of demographic variables thought to impact the magnitude of these differences. Using the WISC-III standardization sample, Prifitera, Weiss, and Saklofske (1998) investigated mean differences on the WISC-III with African Americans, Hispanics, and Whites. Their results indicated that matching participants on age, sex, region of country, parental education level, and number of parents in the household, significantly reduces score

differences between Whites and African Americans and Whites and Hispanics. Consequently, the well-documented 15- point difference between African Americans and Whites (Neisser et al., 1996) becomes a difference of only 11 when the aforementioned variables were used to match the samples. For Hispanics and Whites, the initial 9-point mean score difference decreases to less than 4-points.

Similar results were reported by Naglieri and Ronning (2000) in a study of three matched samples of Whites and African Americans, Whites and Hispanics, and Whites and Asians on the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997). Some researchers have suggested that nonverbal measures provide a favorable alternative to traditional, language loaded assessment methods with culturally and linguistically diverse students (Frisby, 1999; McCallum, Bracken, & Wasserman, 2001) and therefore have the potential to reduce mean score differences. In the Naglieri and Ronning (2000) study, groups were matched on geographic region, socioeconomic status, ethnicity, and type of school setting. Results found significant but small differences between the African American and White samples (mean difference of 4 points) and between Hispanics and Whites (mean difference of 3 points). No significant differences were observed between White and Asian samples. Bracken and McCallum (1998) also found reduced levels of majority and minority differences by matching on age, sex, and highest level of parental education for the UNIT. Differences were reduced to 8.63 and 2.13 for African American and White and Hispanic and White samples on the UNIT Standard Battery FSIQ, respectively.

Presumably, if racial-ethnic group differences are ‘real,’ rather than the result of bias, then attempts to reduce group differences through matched comparison studies

would fail. However, research supports small mean score differences even when controlling for geographic region, socioeconomic status, and other related variables. The process of matching groups on demographic variables does not explain all of the variability in performance among groups. Prifitera et al. (1998) suggests that a more refined matching process on variables that impact equal opportunity to learn and cognitive development may lead to additional reductions in mean score differences. More importantly, identification of factors contributing to lower performance of some groups on cognitive measures has implications for interventions that structure the home and educational environment in a way that maximizes minority student achievement.

Statement of the Problem

The test bias debate often obscures the fundamental finding of large mean score differences between racial and ethnic minority groups and the majority population. Many scholars refute the contention that mean IQ score differences are the result of cultural test bias (Jensen, 1974, 1976, 1980; Reynolds, Willson, & Chatman, 1984; Reynolds, et al., 1999). So, why do these mean score differences exist? According to Puente and Salazar (1998) the goal is to "determine exactly what those differences are, how are they manifested when important variables are controlled, and finally, what do these differences suggest" (p. 244). The first two questions can be addressed by evaluating the extent to which mean differences are effected when minority and majority samples are matched extensively on variables not typically available (to the matching process). That is, will refining the match process, beyond the level typically attained, reduce minority and majority population mean differences? More specifically, will the use of additional matching variables (i.e., community setting and both parental educational levels) further

reduce the minority and majority mean IQ score differences on the UNIT, a measure of nonverbal intelligence, beyond the reduction already found via the use of age, sex, and highest level of one parent's education?

Research Questions

The following research questions are addressed by this study:

1. Are there significant mean score differences between African American and White FSIQs on the UNIT Abbreviated Battery when the following matching variables are used: age, sex, community setting, and both parent education levels?
2. Are there significant mean score differences between African American and White FSIQs on the UNIT Standard Battery when the following matching variables are used: age, sex, community setting, and both parent education levels?
3. Are there significant mean score differences between African American and White FSIQs on the UNIT Extended Battery when the following matching variables are used: age, sex, community setting, and both parent education levels?
4. Are there significant mean score differences between Hispanic and Non Hispanic FSIQs on the UNIT Abbreviated Battery when the following matching variables are used: age, sex, community setting, and both parent education levels?
5. Are there significant mean score differences between Hispanic and Non Hispanic FSIQs on the UNIT Standard Battery when the following matching variables are used: age, sex, community setting, and both parent education levels?
6. Are there significant mean score differences between Hispanic and Non Hispanic FSIQs on the UNIT Extended Battery when the following matching variables are used: age, sex, community setting, and both parent education levels?

CHAPTER 2

Methods

Participants and Procedures

Data for this study were obtained using a subsample of the standardization sample of the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998) provided by Riverside Publishing Company (Appendix C) and additional data collected from UNIT validity studies. The purpose and procedures regarding the study met with full IRB approval (Appendix B). The entire standardization and validity sample used for this study included 3,861 children and adolescents ages 5 years, 0 months, through 17 years, 11 months. The stratified, random sampling procedure, during development of the UNIT, resulted in a sample that closely approximated the U. S. population according to the 1995 U.S. Census data. The following variables were adequately represented in the standardization: sex, race, Hispanic Origin, region, community setting, classroom placement, special education exceptionality, and parental educational attainment. The community setting variable identifies the primary residence of individuals as being in rural, urban, or suburban settings. The U.S. Census defines rural as a community of less than 2,500 people. Parent education, used as an index of SES, was divided into four levels: Less than high school, high school graduate, 1-3 years of college or technical training, and four or more years of college or technical training.

Initial matched comparison studies with White, African American, and Hispanic samples reported in the UNIT manual (Manual) only matched groups on sex, age, ethnicity, and the highest parent education level. Results of these studies yielded African American and White mean FSIQ score differences of 7.63, 8.63, and 9.77 on the

Abbreviated, Standard, and Extended Batteries, respectively (Bracken & McCallum, 1998). The Hispanic and Non Hispanic FSIQ score differences were 2.0, 2.13, and 1.43 on the Abbreviated, Standard, and Extended Batteries, respectively. The current study builds upon these results and examine the African American and White as well as Hispanic and Non Hispanic mean score differences following a more refined match process, specifically by matching groups on two additional variables (community setting and both parent education levels). A total of 168 demographically matched pairs of African Americans and Whites and 162 demographically matched pairs of Hispanics and Non Hispanics were used for analysis. It is important to note that Hispanic designation is based on ethnic origin, of which members of any race may be included. The Non Hispanic sample consists of African Americans, Whites, and/or Asians not of Hispanic origin.

Instrument

The UNIT is an individually administered, multidimensional measure of intelligence designed primarily to provide a fair, accurate, and comprehensive assessment of person's ages 5 years to 17 years. This test is administered in a 100% nonverbal fashion relying entirely on gestures and pantomimed movements as the primary source of communication in the testing situation. McCallum, Bracken, and Wasserman (2001) define "nonverbal assessment" as essentially the assessment of intelligence administered in a nonverbal fashion. Unlike most other nonverbal tests including the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997) and the Test of Nonverbal Intelligence-Third Edition (TONI-III; Brown, Sherbenou, & Johnsen, 1997), the UNIT provides a more comprehensive assessment of intelligence.

The UNIT consists of six subtests, with a mean of 10 and a standard deviation of 3, and provides three administration options. The Abbreviated Battery includes only two subtests and may be used for screening purposes. The Standard Battery consists of four subtests and is typically used for making placement decisions (Bracken & McCallum, 1998). Finally, an Extended Battery option is available for a more comprehensive diagnostic assessment of intelligence. The UNIT assesses a broad array of complex memory and reasoning abilities including those conducive to verbal mediation, using language related symbols (Symbolic) as well as those that are relatively free of symbols and instead require abstractions (Nonsymbolic). Distinctions drawn between symbolic and nonsymbolic content liken that of verbal and performance content, respectively, on the Wechsler scales. The UNIT provides scores on five scales, all with an average standard score of 100 and a standard deviation of 15. These scales include the Memory Quotient, Reasoning Quotient, Symbolic Quotient, Nonsymbolic Quotient, and Full Scale Intelligence Quotient.

The UNIT strongly correlates with traditional language loaded intelligence tests, indicating that the construct measured with this test, is in fact general intelligence (Bracken & McCallum, 1998). Correlational studies reported in the UNIT manual with the Wechsler Intelligence Scale for Children- Third Edition (Wechsler, 1991), Woodcock-Johnson Tests of Cognitive Ability-Revised (1989), and the Matrix Analogies Tests (MAT; Naglieri, 1985) all yield correlational coefficients greater than .81 for the standard and extended batteries. Correlations between the UNIT Abbreviated Battery full scale IQ and the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990), also a screening device, yielded a coefficient of .71. The K-BIT correlated .82 and .79

with the UNIT on the Standard and Extended Batteries, respectively. The UNIT's strong correlation with commonly used full batteries and screening devices support its foundation as a valid and useful measure of intellectual functioning.

The chief goal in developing the UNIT was to provide an accurate assessment of intellectual functioning for individuals with cultural or language related differences that may confound results of intellectual assessment. The authors suggest that the UNIT is especially appropriate for individuals having speech, language or hearing impairments, different cultural and language backgrounds, including those having limited English proficiency. For these groups "traditional, language-loaded intelligence tests do not provide an accurate representation of the true level of intellectual functioning" (McCallum et al., 2001, p. 4). The UNIT was designed to reduce bias associated with the influence of language related demands in the context of intellectual assessment by removing the language component altogether (McCallum et al., 2001), thus providing a more fair assessment of intelligence.

Data Analysis

In this study, a sample of 168 African American and White matched pairs and 162 Hispanics and Non Hispanic pairs were selected from the standardization sample and from subsequent validity studies. Relevant cases in the subsample were extracted for analysis using a statistical package for the social sciences (SPSS). First, a new variable was created using the concatenation procedure in SPSS, which formed a link of the variables of interest: age, sex, community setting, mothers education, and father's education. Next, separate files were created for the samples including African American, White, Hispanic, and Non Hispanic. These samples were then sorted by the new

matching variable, and samples of interest were merged and analyzed. The merge feature in SPSS allowed for the systematic match of demographically identical participants, differing only on the race or ethnicity variable. Data were analyzed using a correlated samples t-test for matched samples. This process was completed for each refinement of the match process. Specifically, data were also analyzed with demographically matched samples on age, sex, community setting, and the highest parent education level.

There are a number of approaches to matching demographically identical participants, particularly when working with large sample sizes. Following the data analysis procedure described above, two additional matching procedures were used to determine the reliability of the results obtained. In the present study, there were significantly more White and Non Hispanic participants in the database than minority individuals. Thus, in many cases, several potential White or Non Hispanic matches were possible for only one minority participant with identical demographics. The two additional procedures required that African Americans and Hispanics be matched with demographically identical White or Non Hispanic participants by (a) using a random numbers table to select one White or Non Hispanic when multiple candidates were available and (b) by matching the first African American and Hispanic with the first identical White or Non Hispanic available. Using both the randomized match and selecting the first demographic match available resulted in mean differences consistent with those obtained using the aforementioned SPSS generated match (see tables 5 and 6; all tables are located in Appendix A), which suggests that the matching procedure used in this study was effective and can be accepted with confidence.

CHAPTER 3

Results

This study examined the effects of increasingly precise matching of socioeconomic variables on mean score IQ differences in nonverbal intelligence test scores. Table 1 (all tables are located in Appendix A) displays the UNIT FSIQ means, standard deviations, and effect sizes for the African American and White samples on the Abbreviated, Standard, and Extended Batteries, as well as mean scores for the Memory Quotient, Reasoning Quotient, Symbolic Quotient, and Nonsymbolic Quotient subscales.

Table 2 displays the UNIT FSIQ means, standard deviations, and effect sizes for the Hispanic and Non Hispanic samples on the Abbreviated, Standard, and Extended Batteries, as well as means for the Memory Quotient, Reasoning Quotient, Symbolic Quotient, and Nonsymbolic Quotient subscales. Several specific research questions were formulated. Results for each research question are discussed in detail. To avoid inflated Type I errors due to the multiple comparisons in the Abbreviated, Standard, Extended batteries, t-test results were adjusted using the Bonferroni technique. Cohens *d* was used to calculate effect sizes. To determine Cohens *d*, the difference between the two group means were divided by the average standard deviation of the two groups. A *d* of .80 or greater is considered large, a *d* of .50 is considered moderate, and a *d* of .20 is small (Cohen, 1988).

Table 3 and table 4 summarize the means, standard deviations, and difference scores of African Americans, Hispanics, and Whites for each separate match process analyzed.

Research Question One

Examination of mean score differences between African Americans and Whites on the UNIT Abbreviated Battery

To determine whether significant mean score differences exist between African Americans and Whites when matched on age, sex, community setting, and both parents education levels, a correlated t-test was run (Table 1). The mean difference of 7.30 between the two groups was statistically significant, t (167) = 4.49, p<.01. A moderate effect size of .48 was obtained according to Cohen's (1988) criteria (for determining the magnitude of effect size). Results suggest that increased precision on socioeconomic variable matching does not further reduce mean score differences between the two groups on the UNIT Abbreviated Battery from initial matched comparison estimates reported in the UNIT manual (mean difference of 7.63). Only negligible reductions were found by equating the two groups on community setting and both levels of parent educational attainment.

An additional analysis equated African Americans and Whites on age, sex, community setting and the higher of two parent education levels These results are reported in Table 3 and denoted as Manual/Community Setting. Table 3 also includes data reported by the UNIT authors as published in the UNIT manual. The label *Manual* will be used to denote this set of matching variables in both Tables 3 and 4. The UNIT authors matched African Americans and Whites on age, sex, and the highest parent's education level. Table 3 allows a comparison of these three separate matching criteria and displays the means, standard deviations, and difference scores based on all three matching processes (e.g., Manual, Manual/community setting, Manual/community

setting/second parent education). On the Abbreviated battery, the difference scores were similar with each match process. Increasing socioeconomic control did not lead to additional reductions in mean IQ score differences.

Research Question Two

Examination of mean score differences between African Americans and Whites on the UNIT Standard Battery

Comparison of the African American and White samples on the UNIT Standard battery resulted in a mean score difference of 8.51. The mean score difference between these two groups was statistically significant, t (167) = 5.17, $p<.01$, when matched on age, sex, community setting, and both parent education levels (Table 1). The mean score differences favored the White sample, and results are similar to those obtained with demographically matched African American and White samples in the UNIT manual (mean difference of 8.63). This difference is smaller than the 15-point difference reported by Neisser et al. (1996) as typical of conventional verbally loaded intelligence tests and lower than the WISC-III group differences even after matching on SES (mean difference of 11). Results suggest that additional refinement on socioeconomic related variables does little to further reduce mean score IQ differences in UNIT Standard FSIQ scores.

Table 3 shows the results of applying all three matching criteria. Data reported by the UNIT authors found a mean difference of 8.63 when equating African Americans and Whites on age, sex, and the higher of two parent education levels. Adding an additional variable, community setting, reduced mean score differences to 8.25, the lowest of the three comparisons. Examination of the FSIQ means (based on the three separate match

processes) shows little change in mean score differences as a function of increased socioeconomic control.

Research Question Three

Examination of mean score differences between African Americans and Whites on the UNIT Extended Battery

Table 1 displays the means, standard deviations, and mean difference scores of the African American and White samples on the UNIT Extended Battery. The mean score difference of 8.72 was statistically significant, t (167) = 5.26, p<.01, in favor of the White sample. The mean difference from this analysis is slightly reduced compared to mean differences reported on matched African American and White samples reported in the UNIT manual (9.77). Thus, additional refinement socioeconomic control does appear to lead to small reductions mean score differences on the extended battery, which is the most comprehensive assessment of general intelligence available from the UNIT administration. However, it is important to note that the decrease in mean score differences resulted from a decrease in the scores of Whites, while the mean FSIQ score for African Americans remained relatively unchanged.

Table 3 displays the means, standard deviations, and difference scores based on all three separate match processes. Equating African Americans and Whites on age, sex, community setting, and the higher of two parent education levels led to mean score differences of 8.87 on the UNIT Extended battery. All analyses resulted in similar difference; however, use of additional matching criteria resulted in slight reductions in mean score differences, compared to studies reported in the UNIT manual. Although overall mean score differences were reduced by including the additional parent education

and community setting variables, inspection of the means reveal that the actual FSIQ means for African Americans were lower than the original Manual match as were means for the White sample.

Research Question Four

Examination of mean score differences between Hispanics and Non Hispanics on the UNIT Abbreviated Battery

To determine whether significant mean score differences exist between Hispanics and Non Hispanics, a correlated t-test was calculated (Table 2). The mean difference of .35 between the two groups was not statistically significant, t (161) = .21, p>.01, indicating that Hispanics and Non Hispanics performed equally as well when matched on age, sex, community setting, and both parents educational levels. The effect size of .02 is considered small according to Cohen's (1988) criteria for determining the magnitude of effect sizes. Results suggest that increased precision on socioeconomic matching does further reduce mean score differences between the Hispanics and Non Hispanics on the UNIT Abbreviated Battery from initial matched comparison estimates reported in the UNIT manual of 2.0.

Table 4 summarizes the means, standard deviations, and difference scores based on the three separate matching processes (e.g., Manual, Manual/community setting, Manual/community setting/second parent education). On the Abbreviated battery, the difference scores were reduced considerably from original Manual differences (2.0) by matching the participants on community setting and both parents educational levels. Increasing socioeconomic control does appear to further reduce mean IQ score differences.

Research Question Five

Examination of mean score differences between Hispanics and Non Hispanics on the UNIT Standard Battery

Comparison of the Hispanic and Non Hispanic samples on the UNIT Standard Battery resulted in a mean score IQ difference of .47. The mean score difference between these two groups was not statistically significant, t (161) = .29, p>.01, when matched on age, sex, community setting, and both parents education level (Table 2). Although the mean score differences favored the White sample, results are smaller than mean score difference results of 2.13 obtained with Hispanics and a demographically matched comparison sample in the UNIT manual. Additional refinement on socioeconomic matching variables does further reduce mean score nonverbal IQ score differences from those typically reported in the literature. Mean score differences favored Hispanic examinees on the Memory Quotient (D = -1.07) and Nonsymbolic Quotient (D = -1.91) subscales.

Table 4 summarizes the means, standard deviations, and difference scores of the UNIT Standard battery based on all three separate match processes. Equating Hispanics and Non Hispanics on age, sex, community setting, and the higher of two parent education levels yielded a mean score difference of .51. Matching these groups on age, sex, community setting, and including both parent education levels in the analysis, yielded mean score differences of .47. These findings, compared to the matched comparison study reported in the UNIT manual, show large reductions in mean score differences as a function of increased socioeconomic control.

Research Question Six

Examination of mean score differences between Hispanics and Non Hispanics on the UNIT Extended Battery

Table 2 displays means, standard deviations, and mean differences of the Hispanic and Non Hispanic samples on the UNIT Extended Battery. The mean score difference of .00 was not statistically significant, t (161) = .000, p>.01, indicating that Hispanics and Non Hispanics earned identical scores when matched on age, sex, community setting, and both parents educational level. These differences are reduced from the difference of 1.43, obtained with demographically matched Hispanic and Non Hispanic samples reported in the UNIT manual. Consequently, additional refinement with regard to socioeconomic status does appear to lead to further reductions in mean score IQ differences. Performance on the Memory Quotient (D=-.74) and Nonsymbolic Quotient (-2.40), favored Hispanic examinees. Thus, on the memory tasks and tasks not mediated by verbal processes, Hispanics outperformed their Non Hispanic counterparts.

Table 4 summarizes the means, standard deviations, and difference scores of the UNIT Extended battery based on the three separate match processes. An analysis equating Hispanics and Non Hispanics on age, sex, community setting, and the higher of two parent education levels yielded a mean score difference of .18. Matching these groups on age, sex, community setting, and both parent education levels, led to a mean score difference was .00. The matched comparison study in the UNIT manual yields a mean score difference of 1.43. With each subsequent match process, the mean score differences are reduced as a function of increased socioeconomic control.

CHAPTER 4

Discussion

The use of intelligence testing in schools has met with significant opposition over the past four decades. Many scholars question the ability of an intelligence test to adequately assess a construct as complex as intelligence. Throughout history, a number of theoretical perspectives have been offered to help explain the form and nature of "intelligence." Some theorists emphasize a dominant, general intelligence construct (Spearman, 1927) whereas others conceptualize intelligence as consisting of a number of specific and largely independent domains (Gardner, 1983). Many commonly used intelligence tests to date, have been developed using these conceptualizations of intelligence as a theoretical underpinning. Apparent even with the earliest attempts at the measurement of intelligence, from Sir Francis Galton (1869) and his theory of inheritable characteristics and intelligence (Fancher, 1998) to Alfred Binet in 1905 and his development of the first "real" intelligence test (Kaufman, 2000), scholars have primarily relied on intelligence measures as a way of distinguishing between "normal" students and lower performing students needing specialized educational services (French, 1986).

A host of newly developed verbal and nonverbal intelligence measures have moved the field of school psychology in new and exciting directions. However, test misuse, misdiagnosis, and misclassification as well as fiscally dictated school psychology practices, often thwart the best intentions of test use for improving student outcomes. If the history of intelligence testing is any indication, then Kaufman (2000) is in all probability correct in his prediction that IQ tests will survive the controversy that has surrounded their existence for decades. The challenge of researchers and practitioners is

to select measures of intelligence that are culturally fair and predictive of academic achievement irrespective of sex, race, ethnicity, or language ability.

Much of the controversy has concerned the use of intelligence tests with culturally diverse minorities. Critics of intelligence tests argue that tests are "biased" against groups who systematically obtain lower IQ scores. Proponents of "test bias" arguments have conceptualized the problem of bias from both the cultural difference (or acculturation) perspective (Helms, 1992) as well as psychometric perspectives, which underscore threats to the tests validity (Hilliard, 1984; Dent, 1996). Although few new tests have serious characteristics of "psychometric bias," they all show mean score differences between culturally and racially diverse group. Those differences continue to raise concerns among examiners who must administer and interpret cognitive ability measures with confidence. These concerns prompted development of the UNIT, a measure built upon the techniques of "fairness."

Mean Score Differences in Demographically Matched Samples

Generally, African American and Hispanic student's score much lower than Whites on verbally loaded measures of intelligence with mean score IQ differences of approximately 15 points and nine points, respectively. Nonverbal intelligence measures, such as the UNIT (Bracken & McCallum, 1998), are designed to reduce the irrelevant influence of culture and language in testing by placing minimal emphasis on expressive and receptive language skills and relying on the use of universally familiar gestures and objects (McCallum, 1999). Matched comparison studies reported in the UNIT manual (Bracken & McCallum, 1998) show relatively small mean score differences between minority and majority groups when equated on age, sex, and highest parent education

level. The purpose of this study was to evaluate the effects of refining the match process by further equating minority and majority groups on age, sex, community setting, and both parent education levels, which suggests increased socioeconomic control.

Given the well documented influence of socioeconomic conditions on intelligence, it is reasonable to expect that careful matching on socioeconomic variables thought to impact the magnitude of this relationship, would greatly reduce mean score group differences. However, results of this study do not unequivocally support this prediction. In fact, the African American and White mean differences do not show a significant reduction as a function of an increasingly precise SES match, in contrast to results obtained by Naglieri and Ronning (2000). These authors found small but significant differences between African American and White samples on a measure of nonverbal ability (mean difference of four points). However, these group differences may have been influenced by the unidimensional nature of the instrument used to examine group differences. The distinction between unidimensional and multidimensional measures of nonverbal intelligence lies in the nature and extent of the skills assessed (Bracken & McCallum, 1998). Unidimensional measures, such as the Naglieri Nonverbal Ability Test (NNAT) assess only narrow skills associated with intelligence by using matrix analogy tasks. On the other hand, multidimensional assessments (e.g., UNIT) take a more comprehensive approach to the assessment of intelligence and provide a better measure of *g*.

Although the present study finds significant differences between African Americans and Whites in UNIT FSIQ scores (with refined matching), these differences are smaller than the one standard deviation typically reported with conventional

intelligence measures. Group differences between African Americans and Whites were similar whether the match process included community setting and one parent's education or both parent education levels. These findings suggest that parental education attainment as an index of socioeconomic status did not add to the understanding of group differences with the African American sample. The fact that precise socioeconomic matching did not produce reductions in mean score IQ differences suggests that the level of refinement necessary was not obtained. Better measures of socioeconomic status should be used to understand the capacity of increased socioeconomic control to reduce mean score IQ differences between these two groups in particular. One potential method to achieving a better SES match, as discussed later in implications, is the control for cross-generational effects. Specifically, matched comparison groups should not only match groups on SES but also control for the length of time within a given socioeconomic stratum.

No significant group differences were found between Hispanics and Non Hispanics in UNIT FSIQ scores on the Abbreviated, Standard, and Extended batteries when equated on age, sex, community setting, and both parent education levels. Thus, mean score differences reported in the UNIT manual were further reduced from 2.0, 2.13, and 1.43 to .35, .47, and .00 on the Abbreviated, Standard, and Extended batteries, respectively. These differences are smaller than those from other measures after matching on SES, such as the three-point differential reported by Naglieri and Ronning (2000) and the approximate four-point difference reported by Prifitera et al. (1998). Hispanic and Non Hispanic mean score differences were smallest when samples were matched on age, sex, community setting, and both parent education levels, as opposed to one parent's

educational attainment. In this case, the more refined the socioeconomic match, the greater the reduction in mean difference scores. In some instances, Hispanics actually outperformed their Non Hispanic counterparts. These findings suggest the UNIT to be a favorable alternative to traditional intelligence measures when assessing language minority students.

Socioeconomic Status as a Predictor of Intellectual Ability

One purpose of the present study was to examine the nature of group differences given increasing control of SES related variability in intelligence test scores. Researchers typically study the relationship between SES and intelligence as a vehicle for understanding how environmental variability works to influence intelligence. The assumption is that those who are higher in SES will have more enriched experiences and therefore higher IQs. Matching on some SES variables does result in reduced mean score differences between racial and cultural groups. However, McLoyd and Ceballo (1998) contend that significant disparities may still exist between ethnic minority and majority groups even when matched on socioeconomic status. Specifically, racial and ethnic minority groups are more likely than Whites to be economically instable, experience more interruptions in income, and lack sufficient savings. African Americans and Hispanics also take on more familial responsibility and contribute more financially to family expenses than do Whites (Goldscheider & Goldscheider, 1991). Ostensibly, even within the same socioeconomic status, comparing minorities and Whites is misleading, as the two groups may still differ significantly at the outset.

Suzuki and Valencia (1997) argue that some socioeconomic indexes, such as parent education, may be more predictive of the intellectual ability for Whites than for

Blacks (White, 1982). The use of parent education as an index of socioeconomic status has been criticized due to the number of variables that work concomitantly to confound this relationship. Often, measures of parent education provide static representations of current economic conditions but offer little information on the socioeconomic history of families, such as the number of generations within a particular socioeconomic stratum. Neisser et al., (1996) suggests that only one generation has passed since the Civil Rights Movement. Researchers should consider the ongoing effects of this relatively recent event and the resulting variation among seemingly homogeneous socioeconomic groups depending upon length of time within the "middle-class." Thus, it is not surprising that the National Center of Educational Statistics (NCES, 2000) found African American and Hispanic students with college educated parents to have significantly lower achievement than White students with parents having college degrees. These cross-generational data are necessary for a more precise match when using parent education as an index of SES.

Comparative Study and Future Directions

Much research has been devoted to the identification and explanation of group differences in cognitive ability and academic achievement. Myerson, Rank, Raines, and Schnitzler (1998) found a positive relationship between the number of years in school and cognitive ability test scores. As indicated earlier, schools have the potential to impact intelligence in a number of ways. Many of the skills required to perform well on conventional intelligence tests are learned in school, and as a result, scores obtained on IQ tests may vary as a function of schooling experiences. A popular method used to study group differences has been comparative analysis. However, across group comparisons are complicated due to the fact that minority and majority students, even of similar

socioeconomic levels, may differ markedly in the quality of their public school education (Farley & Frey, 1994). Myerson et al. (1998) suggests that Whites experience more educational benefit from highschool, whereas African Americans are more likely to benefit from college education. Their findings conclude that cognitive differences may result from historical educational disparities between African Americans and Whites at the secondary education level as these gaps are greatly improved as African Americans enter college, which tend to be academic institutions of equitable quality.

Comparative studies, although limited in explaining the true nature of group differences, are not without benefit. Wong and Rowley (2001) caution against the erroneous interpretation of group differences as minority group deficits, without consideration of critical cultural factors. Low-income minority students develop within a socio-cultural context that is unique to most. These students may experience prolonged exposure to dangerous living conditions, poor schooling experiences, poor quality teachers, poorly educated parents, poor pre- and post-natal care, poverty, low expectations, language barriers, immigration, and historic discrimination and racism. Children who grow and develop within these depressed conditions are less able to profit optimally from their home or school environment, as most children do (Sternberg et al., 2002). The interplay among these socio-cultural factors on intelligence are often lost in simply comparing minority students with majority groups who experience very little to none of these dynamics and are therefore different at the outset. Given these complexities, it is difficult to demonstrate that observed group differences in intelligence reflect inherent capabilities, environmental stimulation, or simply the accumulation of

these historic, pervasive, and harmful risk factors that have been found to play a part in ethnic group differences in IQ (Brooks-Gunn, Klebanov, & Duncan, 1996).

Limitations and Implications for Future Research

A significant limitation of this study lies in the use of parent education as an indicator of socioeconomic status. The parent education variable did not help to explain group differences in the African American and White samples; however, it appeared to further reduce mean score differences within the Hispanic sample. Results of this study were limited to the use and analysis of existing data from the standardization sample of the UNIT. Implications for future research regarding the use of carefully controlled socioeconomic indicators are discussed.

The findings of this study are consistent with the notion that SES is predictive of intellectual performance (also see Oakland, 1978); however, socioeconomic status alone cannot explain all of the variance between minority and majority mean IQ score differences. As the population of American schools becomes increasingly more diverse, so does the need for fair and culturally appropriate intellectual assessment practices. Research using demographically matched samples of minority and majority groups show greater reductions in mean score differences on nonverbal measures (Bracken & McCallum, 1998; Naglieri & Ronning, 2000) as opposed to conventional verbally loaded measures (Prifitera et al., 1998). In the present study, a decrease in the mean score IQ difference was observed when Hispanics and Non Hispanics were matched on relevant socioeconomic variables. These reductions are significant statistically as well as practically since many diagnostic decisions are based on intelligence test results, in which one or two points may make the difference in categorical placements. Future research

should explore nonverbal test use with minority students to determine the potential benefit of these measures in predicting minority student achievement beyond that of traditional verbally loaded measures.

Second, parent educational attainment and community setting are limited in explaining group differences on cognitive ability tests for some populations, relative to others. Decades of research has provided insights into important correlates of intellectual performance of minorities outside of these SES influences including home resources (Gottfried, 1984), educational encouragement (Bahr & Leigh, 1978), schooling and education (Neisser et al., 1996), school composition and neighborhood influences (Brooks-Gunn et al., 1993), parent talk (Hart & Risley, 1995), and family size (Grissmer, Kirby, Berends, & Williamson, 1994) to name a few.

In a meta-analysis examining the relationship between SES and achievement, White (1982) suggests that measures of home environment are correlated to a much higher degree with achievement than any single or combination of SES indexes. These variables may prove to be more important for African American achievement. However, few studies address the substantial variability with regard to the aforementioned influences, across populations, even within the same SES levels. Similar levels on a given socioeconomic indicator (i.e., income, occupation, and living conditions), does not necessarily suggest adequate control for significant environmental influences (Trotman, 1978). In-depth examination of factors that led to reductions in means score differences between the Hispanic and Non Hispanic sample but only small reductions between the African American and White samples (given increased socioeconomic precision) is a critical next step. Determining the extent to which community setting (urban, rural,

suburban) differentially predicts intelligent behavior among Hispanics and African Americans may provide important answers; and requires further study.

Future research should expand the study of homogeneous socio-economic groups to include controls for generational effects that mediate the relationship between income and intelligence. Cross-generational data may improve the predictive relationship of these two variables. If one accepts the assumption that those higher in SES will have more enriched experiences and therefore higher IQs, it may also be argued that those in the higher socioeconomic stratum for longer periods of time will have gained more benefits (e.g., intelligence) from the high SES designation.

Future research on within-ethnic group variables can serve as an impetus to greater understanding of the underlying processes that lead individuals of the same race to experience different outcomes (Wong & Rowley, 2001). The continual use of deficit models, which compare minority and majority groups to explain group differences in IQ scores, perpetuate a circular argument regarding the nature of group differences; but, offer little information about within-group characteristics that may impact this relationship. Further research is needed to explore within racial-ethnic group differences and factors that contribute to or inhibit cognitive growth and academic achievement (Meece & Kurtz-Costes, 2001) in minorities.

The nature of group differences in intelligence is a complex and multifaceted issue and there may be multiple poorly understood sources of this variation. However, what is known is that intelligence tests do not adequately assess all aspects of intelligence (Neisser et al., 1996). Sternberg et al. (2002) notes the difficulty in using and interpreting intelligence tests the same way among different groups. To date, many commonly used

intelligence tests, measure only those abilities valued by mainstream majority population (e.g., logic, reasoning) and have largely disregarded other important aspects of intelligence including divergent thinking and creative problem solving. Unilateral assessment and interpretation procedures may fail to fully tap the abilities of cultural and language minorities. Perhaps it may be beneficial to reexamine other theoretical conceptualizations of intelligence and their utility with diverse populations. In a study comparing general versus specific abilities in achievement of African Americans, Caucasians, and Hispanics, Keith (1999) concluded that a more comprehensive understanding of student's skills may be gained through examination of specific, nontraditional abilities rather than simple assessment of general intelligence, as conventionally defined. Sternberg and Clinkenbeard (1995) suggest that practical intelligence, as often demonstrated in culturally diverse and low-income students, is just as predictive of achievement as conventional tests.

Braden (1999) addresses the extent to which performance assessment can lead to reductions in minority and majority achievement differences. He notes that performance assessment has the potential to decrease performance gaps, due to their focus on authentic problem solving and diversity as well as permitted creativity in responding. Notwithstanding their benefits, performance assessments require reasoning with prior knowledge, increased language demands, and are arguably more complex than traditional measures, which may also make their use considerably more difficult for minority students (Braden, 1999). Current studies on this subject are equivocal and future research should examine the extent to which performance assessments can be used as a

supplement to current intelligence measures, to gain a more comprehensive picture of the student and their abilities.

REFERENCES

References

Arlitt, A. H. (1921). On the need for caution in establishing race norms. *Journal of Applied Psychology, 5,*179-183.

Bahr, S. J., & Leigh, G. K. (1978). Family size, intelligence, and expected education. *Journal of Marriage and the Family, 40,* 331-335.

Bouchard, T. J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212,* 1056.

Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test examiner's manual*. Itasca, IL: Riverside.

Braden, J. P. (1999). Performance assessment and diversity. *School Psychology Quarterly, 14,* 304-326.

Brooks-Gunn, J., Duncan G. J., Klebanov, P. K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology, 99,* 353-395.

Brooks-Gunn, J., Klebanov, P. K., & Duncan, G. J. (1996). Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal characteristics. *Child Development, 67,* 396-408.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of Nonverbal Intelligence – Third Edition.* Austin, TX: ProEd.

Camilli, G., & Shepard, L. A. (1987). *Methods for identifying biased test items.* Thousand Oaks, CA: Sage.

Carroll, J. B. (1993). *Human cognitive abilities.* New York: Cambridge University Press.

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

Mifflin.

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist, 52,* 1051-1058.

Cleary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist, 30 (1),* 15-41.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Copeland, M. L., Conrad, C., & Chansky, N. M. (1978). Validity of two intelligence tests. *Psychological Reports, 42,* 662.

Dana, R. H. (1993). *Multicultural assessment perspectives for professional psychology*. Needham Heights, MA: Allyn & Bacon.

Darling-Hammond, L. (1985). *Equity and excellence: The education status of Black Americans.* New York, NY: College Examination Board.

Dunn, L. M. (1968). Special education for the mildly retarded: Is much of it justifiable? *Exceptional Children, 23(1),* 5-21.

Fan, X., Willson, V. T., & Kapes, J. T. (1996). Ethnic group representation in test construction samples and test bias: The standardization fallacy revisited. *Educational and Psychological Measurement, 56,* 365-381.

Fancher, R. E. (1998). Biography and psychodynamic theory: Some lessons from the life of Francis Galton. *History of Psychology, 1,* 99-115.

Farley, R., & Frey, W. H. (1994). Changes in the segregation of Whites from Blacks during the 1980s: Small steps toward a more integrated society. *American Sociological Review, 59,* 23-45.

Flanagan, D. P., & Ortiz, S. O. (2002). Best practices in intellectual assessment: Future directions. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-IV* (pp. 1351-1372). Bethesda, MD: National Association of School Psychologists.

Flaugher, R. L. (1978). The many definitions of test bias. *American Psychologist, 33,* 671-679.

Flynn, J. R. (1991). *Asian Americans: Achievement beyond IQ.* Hillsdale, NJ: Erlbaum.

French, L. A. (1986). MR testing and evaluation: New dimensions and old concerns. *Psychology in the Schools, 23,* 64-76.

Frisby, C. L. (1999). Straight talk about cognitive assessment and diversity. *School Psychology Quarterly, 14,* 195-207.

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences.* London: Macmillan.

Gardner, H. (1983*). Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Goldscheider, F. K., & Goldscheider, C. (1991). The intergenerational flow of income: Family structure and the status of Black Americans. *Journal of Marriage and the Family, 53,* 499-508.

Gottfried, A. W. (Ed.). (1984). *Home environment and early cognitive development: Longitudinal research.* New York: Academic Press.

Graham, S., & Harris, K. R. (1989). The relevance of IQ in the determination of learning disabilities: Abandoning scores as decision makers. *Journal of Learning Disabilities, 22, 500-503.*

Grissmer, D. W., Kirby, S. N., Berends, M., & Williamson, S. (1994). *Student achievement and the changing American family.* Santa Monica, CA: RAND Corp.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Harnqvist, K. (1968). Relative changes in IQ from 13-18: II. Results. *Scandinavian Journal of Psychology, 9,* 65-82.

Harrington, G. M. (1975). Intelligence tests may favor the majority groups in a population. *Nature, 258*, 708-709.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H. Brooks.

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47,* 1083-1101.

Helms, J. E. (2002). A remedy for the Black-White test score disparity. *American Psychologist, 57,* 303-305.

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

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

Hilliard, A. (1983). Psychological factors associated with language in the education of the African American child. *Journal of Negro Education, 52,* 24-34.

Hilliard, A. (1989). Back to Binet: The case against the use of IQ tests in the schools. *Diagnostique, 14*, 125-135.

Hilliard, A. (2002). Language, culture, and the assessment of African-American

children. In L. Delpit & J. K. Dowdy (Eds.), *The skin that we speak: Thoughts on language and culture in the classroom.* New York: The New Press.

Horn, J. L. (1991). Measurement of intellectual capabilities: Gf-Gc theory. In K. S. McGrew, J. K. Werder, & R. W. *Woodcock, WJ-R Technical Manual* (pp. 197-232). Itasca, IL: Riverside Publishing.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities. Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-93). New York: Guilford.

Jensen, A. R. (1974). How biased are cultural loaded tests? *Genetic Psychology Monographs, 90,* 185-224.

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

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

Jensen, A. R. (1981). *Straight talk about mental tests*. New York, NY: The Free Press.

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

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence.* Needham Heights, MA: Allyn & Bacon.

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

Kaufman, A. S. (2000). Intelligence tests and school psychology: Predicting the future by studying the past. *Psychology in the Schools, 37,* 7-16.

Kaufman, J., Hallahan, D., & Ford, D. Y. (Eds.). (1998). Disproportional representation of minority students in special education. *The Journal of Special Education, 32,* 3-61.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test manual*. Circle Pines, MN: American Guidance Service.

Keith, T. Z. (1999). Effects of general and specific abilities on student achievement: Similarities and differences across ethnic groups. *School Psychology Quarterly, 14,* 239-262.

McCallum, R. S. (1999). A "baker's dozen" criteria for evaluation fairness in nonverbal testing. *School Psychologist, 53,* 40-60.

McCallum, S., Bracken, B., & Wasserman, J. (2001). *Essentials of nonverbal assessment.* New York, NY: Wiley.

McDevitt, T. M., & Ormrod, J. E. (2002). *Child development and education.* Upper Saddle River, NJ: Merrill Prentice Hall.

McLoyd, V. C., & Ceballo, R. (1998). Conceptualizing and assessing economic context; Issues in the study of race and child development. In V. C. McLoyd & L. Steinberg (Eds.), *Studying minority adolescents: Conceptual, methodological, and theoretical issues* (pp. 251-278). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Meece, J. L., & Kurtz-Costes, B. (2001). Introduction: The schooling of ethnic minority children and youth. *Educational Psychologist, 36,* 1-7.

Myerson, J., Rank, M. R., Raines, F. Q., & Schnitzler, M. A. (1998). Race and general cognitive ability: The myth of diminishing returns to education. *American Psychological Society, 9,* 139-142.

Naglieri, J. A. (1985). *Matrix Analogies Test-Expanded Form: Examiner's manual.* San Antonio, TX: The Psychological Corporation.

Naglieri, J. A. (1997). *Naglieri Nonverbal Ability Test*. San Antonio, TX: Psychological Corporation.

Naglieri, J. A., & Ronning, M. E. (2000). Comparison of White, African American, Hispanic, and Asian children on the Naglieri Nonverbal Ability Test. *Psychological Assessment, 12,* 328-334.

Nandakumar, R., Glutting, J. J., & Oakland, T. (1993). Mantel-Haenszel methodology for detecting item bias: An introduction and example using the guide to the assessment of test session behavior. *Journal of Psychoeducational Assessment, 11*, 108-119.

National Center of Educational Statistics. (2000). *National Assessment of Educational Progress (NAEP), 1999 long term trend assessment.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77-101.

Oakes, J., Gamoran, A., & Page, R. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In P. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570-608). New York, NY: Macmillan.

Oakland, T. (1978). Predictive validity of readiness tests for middle and lower socioeconomic status Anglo, Black and Mexican American children. *Journal of Educational Psychology, 70,* 574-582.

Ogbu, J. U. (1994). From cultural differences to differences in cultural frames of

reference. In P. M. Greenfield & R. R. Cocking (Eds.), *Cross-cultural roots of minority child development* (pp. 365-391). Hillsdale, NJ: Erlbaum.

Oswald, D. P., Coutinho, M. J., & Best, A. M. (2000). Community and school predictors of over-representation of minority children in special education. U.S. Department of Education (Grant No. R308FG70020).

Prifitera, A., Weiss, L. G., & Saklofske, D. H. (1998). The WISC-III in context. In A. Prifitera & D. H. Saklofske (Eds.), *WISC-III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 1-38). San Diego, CA: Academic Press.

Puente, A. E., & Salazar, G. D. (1998). Assessment of minority and culturally diverse children. In A. Prifitera & D. H. Saklofske (Eds.), *WISC-III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 227-248). San Diego, CA: Academic Press.

Reschly, D. J. (1981). Psychological testing in educational classification and placement. *American Psychologist, 36,* 1094-1102.

Reynolds, C. R. (1980). Differential construct validity of intelligence as popularly measured: Correlation of age and raw scores on the WISC-R for Blacks, Whites, males, and females. *Intelligence: A Multidisciplinary Journal, 4,* 371-379.

Reynolds, C. R. (1982). Construct and predictive bias. In R. A. Berk (Ed.), *Handbook of methods for detecting bias*. Baltimore, MD: Johns Hopkins University Press.

Reynolds, C. R., Lowe, P. A., & Saenz, A. L. (1999). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *Handbook of school psychology*- 3rd Ed (pp. 549-595). New York: Wiley.

Reynolds, C. R., Willson, V. L., & Chatman, V. L. (1984). Item bias on the 1981

revisions of the Peabody Picture Vocabulary Test using a new method of detecting bias. *Journal of Psychoeducational Assessment, 2,* 219-221.

Sattler, J. M. (1974). *Assessment of children's intelligence.* Philadelphia: Saunders.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.), San Diego: Jerome M. Sattler.

Scarr, S., & Weinberg, R. A. (1976). IQ test performance of black children adopted by white families. *American Psychologist, 31,* 726-739.

Scheuneman, J. D. (1987). An argument opposing Jensen on test bias: The psychological aspects. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 155-170). Philadelphia: The Falmer Press.

Schofield, J. (1991). School desegregation and intergroup relations. *Review of Research in Education, 17,* 335-409.

Spearman, C. (1927). *The abilities of Man*. New York: Macmillan.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J., & Clinkenbeard, P. (1995). A triarchic view of identifying, teaching, and assessing gifted children. *Roeper Review, 17,* 255-260.

Sternberg, R. J., Grigorenko, E. L., Ngorosho, D., Tantufuye, E., Mbise, A., Nokes, C. et al. (2002). Assessing intellectual potential in rural Tanzanian school children. *Intelligence, 30,* 141-162.

Sturn, A., & Johnston, J. (1999). Thinking out loud: An exploration of problem-solving language in preschoolers with and without language impairment. *International Journal of Language and Communication Disorders, 34,* 1-15.

Suzuki, L. A., & Valencia, R. R. (1997). Race-ethnicity and measured intelligence: Educational implications. *American Psychologist, 52,* 1103-1114.

Taylor, O. L., & Lee, D. L. (1995). Standardized tests and African-American children: Communication and language issues. In A. G. Hilliard (Ed.), *Testing African-American students* (2nd ed., pp. 37-50). Chicago, IL: Third World Press.

Thurstone, L. L. (1960). *The nature of intelligence.* Littlefield: Adams.

Torrance, E. P. (1971). Are the Torrance tests of creative thinking biased against or in favor of "disadvantaged" groups? *Gifted Child Quarterly, 15,* 75-80.

Trotman, F. K. (1978). Race, IQ, and rampant misrepresentations: A reply. *Journal of Educational Psychology, 70,* 478-481.

Wagner, R. K. (1997). Intelligence, training, and employment. *American Psychologist, 52*, 1059-1069.

Wechsler, D. (1971). Intelligence: Definition, theory, and the IQ. In R. Cancro (Ed.), *Intelligence: Genetic and environmental influences* (pp. 50-55). New York: Gruene and Stratton.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third edition*. San Antonio, TX: The Psychological Corporation.

Wells, A., & Crain, R. (1994). Perpetuation theory and the long term effects of school desegregation. *Review of Educational Research, 64*, 53-76.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91,* 461-481.

Wijeyesinghe, C. L., Griffen, P., & Love, B. (1997). Racism curriculum design. In M.

Adams, L. A. Bell, & P. Griffin (Eds.), *Teaching for diversity and social justice* (pp. 82-109). New York: Routledge.

Wong, C. A., & Rowley, S. J. (2001). The schooling of ethnic minority children: Commentary. *Educational Psychologist, 36,* 57-66.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psychoeducational Battery-Revised.* Itasca, IL: Riverside.

APPENDICES

APPENDIX A

TABLES

Table 1

UNIT performance by African Americans and a demographically matched White sample

Scale	African American (n=168) Mean	SD	White (n=168) Mean	SD	D	Effect Size	t-value
Abbreviated Battery							
FSIQ	90.69	14.75	98.00	15.13	7.30	.48	4.49*
Standard Battery							
Memory Quotient	91.10	15.09	98.66	15.49	7.56	.49	
Reasoning Quotient	90.08	15.51	97.61	14.88	7.52	.49	
Symbolic Quotient	90.80	15.79	98.37	15.20	7.57	.48	
Nonsymbolic Quotient	90.25	14.41	97.83	15.13	7.57	.51	
FSIQ	89.30	15.39	97.81	15.43	8.51	.55	5.17*
Extended Battery							
Memory Quotient	90.77	15.85	98.84	15.71	8.06	.51	
Reasoning Quotient	89.91	16.33	97.48	14.47	7.57	.49	
Symbolic Quotient	90.51	16.59	98.68	15.52	8.16	.50	
Nonsymbolic Quotient	90.14	14.92	97.60	15.14	7.46	.49	
FSIQ	89.19	16.15	97.91	15.34	8.72	.55	5.26*

Note: Significance after Bonferroni adjustment: * $p<.01$; UNIT= Universal Nonverbal Intelligence Test; D= White mean- African American mean; Samples were matched according to age, sex, community setting, and both parents educational level.

Table 2

UNIT performance by Hispanics and a demographically matched Non Hispanic sample

Scale	Hispanics (n=162) Mean	SD	Non Hispanics (n=162) Mean	SD	D	Effect Size	t-value
Abbreviated Battery							
FSIQ	98.06	13.50	98.41	16.14	.35	.02	.21
Standard Battery							
Memory Quotient	100.3	14.20	99.31	16.06	-1.07	-.07	
Reasoning Quotient	95.25	13.66	97.18	14.20	1.92	.13	
Symbolic Quotient	95.91	14.91	98.79	15.03	2.87	.19	
Nonsymbolic Quotient	99.59	12.89	97.67	15.92	-1.91	-.13	
FSIQ	97.48	13.66	97.95	15.71	.47	.03	.29
Extended Battery							
Memory Quotient	100.0	14.38	99.27	16.57	-.74	-.04	
Reasoning Quotient	97.09	13.19	97.59	15.80	.49	.03	
Symbolic Quotient	96.88	14.74	98.90	16.03	2.02	.13	
Nonsymbolic Quotient	100.3	12.94	97.91	16.13	-2.40	-.16	
FSIQ	98.29	13.96	98.29	16.49	.00	.00	.00

Note: UNIT= Universal Nonverbal Intelligence Test; D= Non Hispanic mean- Hispanic mean; Samples were matched according to age, sex, community setting, and both parents educational level.

Table 3

Summary of African American and White FSIQ scores on Abbreviated, Standard, and Extended Batteries when matched on age, sex, highest parent education level (Manual); age, sex, community setting, highest parent education level; and age, sex, community setting and both parent education levels

	Manual* (n= 352)					Manual/ Community Setting (n= 208)					Manual/Community Setting/Second Parent (n= 168)				
	African American		White			African American		White			African American		White		
Scale	Mean	SD	Mean	SD	D	Mean	SD	Mean	SD	D	Mean	SD	Mean	SD	D
Abbreviated															
FSIQ	91.34	12.57	98.97	13.77	7.63	90.46	14.27	98.11	14.42	7.65	90.69	14.75	98.00	15.13	7.30
Standard															
FSIQ	90.68	12.29	99.31	12.17	8.63	89.19	14.73	97.44	14.41	8.25	89.30	15.39	97.81	15.43	8.51
Extended															
FSIQ	90.15	13.18	99.92	12.10	9.77	89.12	15.61	98.00	14.57	8.87	89.19	16.15	97.91	15.39	8.72

*Source: Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test examiner's manual*. Itasca, IL: Riverside.

Table 4

Summary of Hispanic and Non Hispanic FSIQ scores on Abbreviated, Standard, and Extended Batteries when matched on age, sex, highest parent education level (Manual); age, sex, community setting, highest parent education level; and age, sex, community setting and both parent education levels

	Manual* (n= 194)					Manual/ Community Setting (n= 206)					Manual/Community Setting/Second Parent (n= 162)				
	Hispanic		Non Hispanic			Hispanic		Non Hispanic			Hispanic		Non Hispanic		
Scale	Mean	SD	Mean	SD	D	Mean	SD	Mean	SD	D	Mean	SD	Mean	SD	D
Abbreviated															
FSIQ	97.98	12.65	99.98	13.01	2.0	97.60	13.61	98.87	16.25	.27	98.06	13.50	98.41	16.14	.35
Standard															
FSIQ	98.32	12.73	100.4	12.54	2.13	96.97	13.58	97.49	15.73	.51	97.48	13.66	97.95	15.71	.47
Extended															
FSIQ	99.41	13.15	100.8	12.36	1.43	97.80	14.01	97.99	16.08	.18	98.29	13.96	98.29	16.49	.00

*Source: Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test examiner's manual*. Itasca, IL: Riverside.

Table 5

Means, standard deviations, and difference FSIQ scores based on follow-up matching procedures for African Americans and Whites

	Randomized Matching (n = 146)					First Identical Match (n =146)				
	African Americans		Whites			African Americans		Whites		
Scale	Mean	SD	Mean	SD	D	Mean	SD	Mean	SD	D
Abbreviated	91.64	15.37	99.00	14.14	7.36	91.13	14.88	97.94	15.16	6.80
Standard	89.60	16.23	98.98	14.10	9.37	89.64	15.67	97.69	15.27	8.05
Extended	89.38	17.00	98.45	13.35	9.07	89.39	16.31	97.67	15.30	8.28

Table 6

Means, standard deviations, and difference FSIQ scores based on follow-up matching procedures for Hispanics and Non Hispanics

	Randomized Matching (n =121)					First Identical Match (n =121)				
	Hispanics		Non Hispanics			Hispanics		Non Hispanics		
Scale	Mean	SD	Mean	SD	D	Mean	SD	Mean	SD	D
Abbreviated	98.86	13.20	96.52	16.79	2.34	97.80	13.99	94.82	15.60	2.97
Standard	98.46	13.56	95.97	16.11	2.49	97.18	13.78	94.23	15.56	2.94
Extended	99.28	13.85	96.16	16.95	3.11	98.08	14.23	94.72	16.21	3.35

APPENDIX B

COPY OF ACCEPTED IRB APPLICATION

FORM A APPLICATION

IRB #_____

Certification for Exemption from IRB Review for Research Involving Human Subjects

---

A. PRINCIPAL INVESTIGATOR(s) and/or CO-PI(s):

LaRonta M. Upson

B. DEPARTMENT:

Department of Educational Psychology and Counseling

C. COMPLETE MAILING ADDRESS AND PHONE NUMBER OF PI(s) and CO-PI(s):

LaRonta M. Upson: 712 Chestnut Oak Dr., Knoxville, TN 37909 (HOME);

CA525 Claxton Addition, Department of Educational Psychology and Counseling, Department of Education, Knoxville, TN 37996 (OFFICE)

D. TITLE OF PROJECT:

Effects of an increasingly precise socioeconomic match process on mean score differences in nonverbal intelligence test scores.

E. EXTERNAL FUNDING AGENCY AND ID NUMBER (if applicable):

F. GRANT SUBMISSION DEADLINE (if applicable):

G. STARTING DATE: (NO RESEARCH MAY BE INITIATED UNTIL CERTIFICATION IS GRANTED.)

Upon IRB approval

H. ESTIMATED COMPLETION DATE

August 2003

I. RESEARCH PROJECT:

1. Objective(s) of Project (Use additional page, if needed.):
The objective is to determine the effects of matching groups on demographic variables (age, sex, both parents education level, and community setting) on mean IQ score differences. Standardization data from the Universal Nonverbal Intelligence Test (UNIT) were used to match individuals. Initial matched comparison studies reported in the UNIT manual matched participants on age, sex, and highest parent education level. This matching procedure reduced traditional majority-minority group IQ mean differences (Bracken & McCallum, 1998). The purpose of this study is to evaluate the effects of further refining the match process by equating groups not only on age, sex, and highest parent education level, but two additional variables: community setting and both parent education levels. This analysis will address whether the additional refinement may reduce the mean score differences that exists between groups (i.e., Euro-American vs. African-Americans and Euro-American vs. Hispanic Americans).

2. Subjects (Use additional page, if needed.):

Data for this study will be obtained using a subsample of the standardization sample of the Universal Nonverbal Intelligence Test provided by Riverside Publishing Company. The entire standardization sample included 2,100 children and adolescents ages 5 years, 0 months, through 17 years, 11 months. (See attached permission letter)

3. Methods or Procedures (Use additional page, if needed.):
A sample of African American vs. Whites, and Hispanic Americans vs. Non Hispanics will be matched on age, sex, community setting, and both parent education levels using SPSS program. Additional matching variables (not included in initial matched comparison studies) include a refinement by equating the educational level of both parents (rather than one parent) and community setting. The mean score differences in intelligence test scores between these two groups will be examined following this refined match process. Data will be analyzed using an independent samples t-test.

4. CATEGORY(s) FOR EXEMPT RESEARCH PER 45 CFR 46 (see reverse side for categories): Category 4

J. CERTIFICATION: The research described herein is in compliance with 45 CFR 46.101(b) and presents subjects with no more than minimal risk as defined by applicable regulations.

Principal Investigator_____ _____ _____
LaRonta M. Upson 5/22/03

Student
Advisor_____ _____ _____

R. Steve McCallum 5/22/03

Dept. Review
Comm.Chair_____ _____ _____
Robert L. Williams 6/03/03

APPROVED:
Dept. Head_____ _____ _____

R. Steve McCallum 6/03/03

For additional information on Form A, contact Brenda Lawson by e-mail at blawson@tennessee.edu or by phone at (865) 974-7697.

APPENDIX C

COPY OF PERMISSION FROM RIVERSIDE PUBLISHING COMPANY

February 18, 2003

Dave Madsen, Ph.D.
Riverside Publishing
425 Spring Lake Dr.
Itasca, IL 60143-2079

Dear Dr. Madsen:

My name is LaRonta Upson and I am a doctoral student in the school psychology program at the University of Tennessee, Knoxville.

My purpose for this letter is to acquire your permission to use the standardization data collected on the Universal Nonverbal Intelligence Test (UNIT) for further analysis in a dissertation project. With these data we will refine the SES matching process and determine the effect of this process on mean difference scores (e.g., Euro-Americans vs. African-Americans). I will be working closely with R. Steve McCallum on this project, as he is the chair on my dissertation committee. Consistent with the requirements of our Human Subjects Review Board at the University of Tennessee, I am required to obtain your permission to use the standardization data prior to starting my study.

I hope you find this study useful and grant permission to proceed. If you have any further questions, feel free to contact me at: (865) 212-5504, or by e-mail: Lupson@utk.edu. Thank you very much for your assistance.

Sincerely,

LaRonta Upson
Educational Psychology and Counseling
University of Tennessee, Knoxville

_____

R. Steve McCallum
Department Head and Associate Professor

03/13/03 09:33 FAX 6304676144 RIVERSIDE PUB. 002

Riverside Publishing
A HOUGHTON MIFFLIN COMPANY

Thursday, March 13, 2003

LaRonta Upson
University of Tennessee—Knoxville
College of Education, Health, and Human Services
Educational Psychology and Counseling
Claxton Complex A 525
Knoxville, TN 37996-3452
Fax: 865.974.0135

Dear Ms. Upson,

Thank you for your interest in the *Universal Nonverbal Intelligence Test™ (UNIT™)*. This letter will serve as preliminary permission to use the *UNIT* standardization data in your dissertation project. Kirk Becker will work with you to provide the data you need. Formal written permission from our Contracts/ Permissions Department will follow shortly.

Please send me a copy of your research results upon completion.

Sincerely,

Karyn Pastorino

Karyn M. Pastorino
Senior Marketing Coordinator, Clinical Products and Services
Phone: 630.467.6865
Fax: 630.467.6194
Email: karyn_pastorino@hmco.com

VITA

LaRonta M. Upson is from Jacksonville, Florida where she graduated from Jean Ribault Senior High School in 1994. She received her Bachelor of Science in Psychology from The Florida State University in 1998. In 2001, she received combined Master of Science and Educational Specialist graduate degrees in School Psychology also from The Florida State University. She will be receiving her Ph.D. in Education with a concentration in school psychology following completion of her APA accredited internship with the Louisiana School Psychology Internship Consortium in New Orleans, Louisiana.

During graduate school at The Florida State University, she gained professional experiences as a psychological examiner, career counselor, and instructor in career development. She also completed a 1200-hour internship with the Duval County School System during the 2000-2001 school year. At the University of Tennessee she has worked as a graduate assistant for the Associate Dean, editorial assistant for the *Journal of Psychoeducational Assessment*, graduate teaching associate in the department of Educational Psychology and Counseling, and Beginning Spanish instructor for the Project GRAD Summer Institute sponsored by the Office for Faculty and Student Diversity in Academic Affairs.

She is a member of the Honor Society of Phi Kappa Phi, a Holmes Scholar, the National Honor Society in Psychology (PSI CHI), the National Association of School Psychologists, the American Psychological Association, APA Division 16, Mid-South Educational Research Association, and the American Educational Research Association.