Assessment of Factor Models Underlying the WISC-III
in White, Black, and Hispanic Subgroups of
the Standardization Sample

by

Vincent Pandolfi

Dissertation Committee

8/27/97

Charles A. Dill, Ph.D., Sponsor

Charles F. Levinthal, Ph.D., Chair

Joseph R. Scardapane, Ph.D.

Marc D. Summers, Ph.D., Reader

Junko Tanaka-Matsumi, Ph.D.

Submitted in partial fulfillment of the requirements for the
degree of Doctor of Philosophy

Hofstra University

Hempstead, N.Y. 11550

June 23, 1997

UMI Number: 9804828

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Abstract

Structural and measurement invariance of the WISC-III was examined across White (N= 1542), Black (N= 338), and Hispanic (N= 242) subgroups of the standardization sample. Data analyzed were separate subtest scaled score and raw score variance-covariance matrices for each subgroup. Both sets of scores were analyzed as scaled scores may mask unique response patterns within each subgroup.

Within groups and simultaneous maximum likelihood confirmatory factor analyses were performed to fit data to four competing correlated factor models: (a) a one-factor model consisting of all 13 WISC-III subtests; (b) a verbal-performance factor model; (c) a Verbal Comprehension, Perceptual Organization, and Processing Speed model; and (d) a Verbal Comprehension, Perceptual Organization, Freedom From Distractibility, and Processing Speed model. It was hypothesized that verbal and performance factors would fit the data best for each group. It was further hypothesized that factor loading patterns would differ across groups and that analyses of raw score data would reveal idiosyncratic response patterns across groups.

The chi-square/df ratio, Tucker-Lewis Index, and Adjusted Goodness of Fit Index indicated that the four-factor model fit both sets of data best within each racial-ethnic group. These indices and an incremental fit index demonstrated that the four-factor model exhibited structural

and measurement invariance across groups. The same four factors explained the variance-covariance matrices of each group and WISC-III subtests are measured with the same reliability. Differences in rank order of subtest factor loadings were observed when scaled score data were analyzed which was not expected.

Test development procedures safeguarding against bias and acculturation factors may account for the structural and measurement invariance of the four-factor model. Item content of the WISC-III was meant to appeal to a multicultural society. Geographic proximity and intermixing between racial-ethnic groups may also account for the results. The four-factor model may be used clinically in assessing children from White, Black, and Hispanic groups. Since the groups studied were heterogeneous with respect to cultural practices and socioeconomic status, practitioners should not disregard racial-ethnic group membership when assessing children from diverse backgrounds.

DEDICATION

To Caroline, whose companionship made the last five years the most memorable and satisfying of my life. Your unending love and devotion fulfill me.

ACKNOWLEDGMENTS

Vincent and Margaret Pandolfi instilled the values and skills that allowed me to achieve my most important ambitions. With their love and support I have been able to pursue a most fulfilling life course. Julie, I appreciate all the support and encouragement you provided me especially during the past few years.

I thank all the members of my dissertation committee. Dr. Charles Dill made the whole process a relatively smooth journey by always making himself available for consultations. He allowed for my independence in completing this project while providing sound direction. Dr. Junko Tanaka-Matsumi made several critical suggestions that made this project's methodological novelty an important contribution to the psychological literature. Thanks to Dr. Joseph R. Scardapane for his practical feedback and humor which made the process enjoyable. Thank you to Dr. Marc D. Summers for taking the time to make a valuable contribution as an outside reader. Finally, thanks to Dr. Charles F. Levinthal who agreed to chair my orals committee at the last minute.

Finally, special thanks to The Psychological Corporation for providing me with the data for this study.

Table of Contents

Chapter Page

Copyright..ii

Abstract..iii

Dedication

Acknowledgements

List of Tables...v

List of Figures..vii

List of Appendices..viii

I. Introduction..1

Overview and Rationale for the Study...............1

Assessment of Children From Different Racial-Ethnic Groups..6

Factors Influencing Cognitive Skills Development In Children From Different Racial-Ethnic Groups...10

Wechsler's Definition of Intelligence.............12

Organization of the Children's Scales.............16

Factor Analysis and Wechsler Tests................18

WISC Factor Structure: Standardization Data.......21

Summary of WISC Factor Analytic Research: Standardization Data.............................25

WISC Factor Structure in Different Racial-Ethnic Groups..26

Summary of _WISC_ Factor Analytic Research: Racial-Ethnic Subgroup Data............................30

WISC-R Factor Structure: Unrestricted Analyses of Standardization Data...........................32

Chapter Page

Summary of WISC-R Factor Structure: Unrestricted Analyses of Standardization Data...............38

WISC-R Factor Structure: Restricted Analyses of Standardized Data.............................40

Summary of WISC-R Factor Analytic Research: Standardization Data............................43

WISC-R Factor Structure in Racial-Ethnic Sub-groups...44

Summary of WISC-R Factor Analytic Research: Racial-Ethnic Subgroup Data...................56

Factor Analyses of the WISC-III: Standardization Data..68

WISC-III Factor Analytic Research: Racial-Ethnic Subgroups..................................72

Summary of WISC-III Factor Analytic Research......73

Hypotheses..76

II. Method..83

Data..83

Procedure...84

III. Results......................................89

Within Groups Analyses: General Findings..........89

Scaled Score Data.............................90

Raw Data......................................93

General Findings: Simultaneous Analyses...........95

Scaled Score Data.............................96

Raw Data......................................102

Chapter Page

Subtest Factor Loadings....................................102

Summary of Results...113

IV. Discussion...114

Evaluation of the Hypotheses.....................114

Theoretical Issues...............................117

Wechsler's Conceptualization of Intelligence..117

The Multicultural Literature..................118

Methodological Issues............................122

Within Group Heterogeneity....................122

Statistical Issues............................123

Application of the Present Study's Findings......126

Utility of the Four-Factor Model..............126

Conclusion...128

References...132

Appendices...141

List of Tables

Table Page

1. WISC Standardized Subtest Factor Loadings: Standardization Data....27
2. WISC Standardized Subtest Factor Loadings From White Subgroup Data....31
3. WISC Standardized Subtest Factor Loadings From Black Subgroup Data....33
4. WISC Standardized Subtest Factor Loadings From Hispanic Subgroup Data....34
5. WISC-R Subtest Factor Loadings: Standardization Data....39
6. WISC-R Standardized Subtest Factor Loadings: White Subgroup Data....57
7. WISC-R Standardized Subtest Factor Loadings: Black Subgroup Data....60
8. WISC-R Standardized Subtest Factor Loadings: Hispanic Subgroup Data.... 62
9. WISC-R Factor Structure Based on Literature Review of White, Black, and Hispanic Samples: Two-Factor Solution....64
10. WISC-R Factor Structure Based on Literature Review of White, Black, and Hispanic Samples: Three-Factor Solution....67
11. WISC-III Standardized Factor Loadings: Standardization Sample Data....74

Table Page

12. Within Groups Confirmatory Factor Analyses of Scaled Score Data..91

13. Within Groups Confirmatory Factor Analyses of Raw Data..94

14. Simultaneous Confirmatory Factor Analyses of Scaled Score Data Across White, Black, and Hispanic Subgroups...97

15. Simultaneous Confirmatory Factor Analyses of Raw Data Across White, Black, and Hispanic Subgroups.........103

16. Subtest Unstandardized Factor Loadings for Fully Constrained Four-Factor Model: Raw Data vs. Scaled Score Data...107

17. Subtest Unstandardized Factor Loadings for White Sample Four-Factor Model: Raw Data vs. Scaled Score Data...109

18. Subtest Unstandardized Factor Loadings for Black Sample Four-Factor Model: Raw Data vs. Scaled Score Data.. 110

19. Subtest Unstandardized Factor Loadings for Hispanic Sample Four-Factor Model: Raw Data vs. Scaled Score Data..112

List of Figures

Figure Page

1. One-Factor Model....................................77

2. Two-Factor Model....................................78

3. Three-Factor Model..................................79

4. Four-Factor Model...................................80

5. Fully Constrained Four-Factor Model: Scaled Score Data...98

6. Standardized Factor Loadings For Four-Factor Model: Scaled Score Data..................................99

7. Fully Constrained Four-Factor Model: Raw Score Data...104

8. Standardized Factor Loadings For Four-Factor Model: Raw Score Data....................................105

List of Appendices

Appendix Page

A. Within Groups Confirmatory Factor Analyses for Three-Factor Model....141

B. Simultaneous Confirmatory Factor Analyses of Three-Factor Model....142

C. Fully Constrained Three-Factor Model: Scaled Score Data....143

D. Fully Constrained Three-Factor Model: Raw Score Data....144

E. Subtest Unstandardized Factor Loadings for Fully Constrained Three-Factor Model: Raw Data vs. Scaled Score Data....145

F. Subtest Unstandardized Factor Loadings for White Sample Three-Factor Model: Raw Data vs. Scaled Score Data....146

G. Subtest Unstandardized Factor Loadings for Black Sample Three-Factor Model: Raw Data vs. Scaled Score Data....147

H. Subtest Unstandardized Factor Loadings for Hispanic Sample Three-Factor Model: Raw Data vs. Scaled Score Data....148

I. Subtest Standardized Factor Loadings for Four-Factor Model: Scaled and Raw Score Data....149

J. Factor Intercorrelations For The Four-Factor Model Within Each Group: Scaled and Raw Score Data....150

Chapter I

INTRODUCTION

Overview and Rationale for the Study

The Wechsler Intelligence Scale for Children- Third Edition (WISC-III; Wechsler, 1991) is the latest Wechsler instrument to assess children's intelligence from ages 6-16 years. Like the WISC (Wechsler, 1949) and WISC-R (Wechsler, 1974), the test assesses verbal (Verbal Scale I.Q.), nonverbal (Performance Scale I.Q.), and overall (Full Scale I.Q.) abilities. Also like its predecessors, the WISC-III has generated a wealth of research and investigating the test's construct validity is a prominent area of study. This study investigates the factor structure of the WISC-III across White, Black, and Hispanic subgroups of the standardization sample. Factorial invariance across racial-ethnic subgroups has implications for fairness in psychoeducational assessment and Wechsler's theory of intelligence.

David Wechsler's conceptualization of intelligence evolved from previous empirical and theoretical writings as well as his clinical observations. His intelligence tests were derived from early versions of the Stanford-Binet and Army Alpha tests. By assessing a person's intelligence, one could learn much about his or her personality as well as intellect.

Wechsler defined intelligence as the ability of a person to adapt to his or her environment. Intelligence is multifaceted and can be expressed many different ways. Thus, a person may exhibit many abilities, no one more important than the other. Based on the empirical work of Spearman and W.B. Alexander, he subdivided general intelligence into verbal and nonverbal abilities. Within each of these major areas were several specific abilities. Therefore, not only could a person express general intelligence verbally and nonverbally, but one could also express the verbal and nonverbal abilities in many ways. The ultimate measure of intelligence would be how these abilities allowed a person to adapt to his or her environment.

Implementation of these abilities could be affected by a number of nonintellective factors. Wechsler cited motivation and conative traits as two factors that could influence behavior. The capacity to problem-solve on any given occasion can be jointly determined by one's intellectual ability and, for example, one's motivation to solve the problem. Wechsler's Performance Scale was, in part, an attempt to quantify nonintellective variables.

Within the last 40 years of factor analytic research on the Wechsler scales, a distractibility factor (Freedom From Distractibility) often accounts for some variance. The specific nature of this factor has long been a source of controversy but it seems to reflect one of Wechsler's nonintellective variables. Wechsler's verbal, performance, and distractibility factors (to a lesser extent) have withstood many empirical analyses.

Historically, factor analyses of the Wechsler scales have employed many different methods. Not surprisingly, the nature of the constructs underlying the WISC, WISC-R, and now the WISC-III has not been consistent across studies. Although verbal and performance factors generally emerge, the distractibility factor does not emerge in every study. In addition, the content of these constructs vary across racial-ethnic subgroups. Accurate assessment of minority subgroups would entail more precise delineation of the constructs underlying the WISC-III for these subgroups.

Recently, greater sensitivity to the effect of culturally specific learning experiences on intellectual expression has been noted. Such diversity of experience may foster different intellectual competencies across cultural subgroups. Or, similar competencies may develop with different modes of expression of these abilities. Exploratory factor analytic research of the WISC-R often suggests the latter phenomenon when a two-factor solution is indicated: verbal and performance factors arise across racial-ethnic subgroups, however their specific content may vary. The same is true of three-factor solutions even with the addition of Freedom From Distractibility.

Irvine & Carroll (1981) cited a number of problems inherent within many factor analytic approaches to cross-cultural construct validation. These problems exist in study of the Wechsler tests. Historically, establishing racial-ethnic construct equivalence of Wechsler tests has been methodologically flawed. Studies of racial-ethnic subgroup samples that are not representative of their populations and the use of different factor analytic methods have plagued this line of research.

Additionally, studies have used subtest scaled scores in factor analyses. Scaled scores are based on the standardization sample's collective raw score data. This blends data generated by children of many different racial-ethnic subgroups. Hence, unique response patterns within racial-ethnic subgroups may be masked when scaled score data are used in factor analyses. Even when a study examines factor structure within a specific racial-ethnic subgroup, the unit of data analyzed is a standardized conglomerate of multiracial-ethnic input. Factor analysis of raw score data within different racial-ethnic subgroups may allow for more precise delineation of constructs underlying Wechsler tests for each subgroup.

The practical implication of different construct content is great within the domains of theory and psychological assessment. Clusters of verbal and nonverbal subtests may differ depending on an individual's culturally-specific learning experiences. This would mean that

Wechsler's traditional grouping of specific subtests along verbal and nonverbal lines may need modification. Evaluating the content of an examinee's intellectual abilities may depend on culturally-specific experiences. A more precise discussion of various abilities can help generate more specifically tailored recommendations for remediation. Speaking of WISC-III constructs in the same way for all racial-ethnic subgroups may disregard different learning histories rooted in cultural practices. The result may run counter to ethical and fair testing practice. Until conceptual equivalence of WISC-III constructs is empirically demonstrated, with a methodology that addresses previous shortcomings, cross racial-ethnic comparisons of these constructs is meaningless.

The present study addressed methodological weaknesses of earlier factor analytic studies of Wechsler tests across racial-ethnic subgroups. Confirmatory factor analyses were performed on scaled score and raw score subtest covariance matrices generated by White, Black, and Hispanic subgroups of the standardization sample. The present methodology helped determine if traditional use of scaled scores may have masked unique response patterns within the subgroups. This is especially true for the racial-ethnic minority samples which have fewer subjects and contribute less to the subtest covariance matrix than the White subjects' data. Use of the standardization sample helped ensure sufficient sample size to stabilize factor solutions. In addition, each subgroup was representative of their respective racial-ethnic populations within the United States.

Four empirically-derived factor models purported to underlie the WISC-III were assessed. The analysis determined the extent to which models demonstrated structural and/or measurement invariance across the three subgroups. The study helped clarify conceptual issues regarding WISC-III constructs across racial-ethnic subgroups. Finally, the study helped determine if Wechsler's theory of intelligence can be applied the same way to three racial-ethnic subgroups. It was hypothesized that while two factors (verbal and performance) may explain the covariance structure for each of the subgroups, the pattern of factor loadings will differ.

Assessment of Children From Different Racial-Ethnic Groups

The term "racial-ethnic" will be used throughout this paper rather than "racial," "ethnic," or "cultural." Within the psychological literature, these terms are often used interchangeably. No single definition exists for these terms (Okazaki & Sue, 1995). Since Wechsler (1991) stratified their sample based on "race-ethnicity," this paper will use this term to maintain consistency with their methodology.

The psychometric literature reviewed below illustrates attempts at determining whether the WISC-III measures similar constructs across different racial-ethnic subgroups of the population. Such research has implications for the field of psychological assessment. If the test's constructs may be interpreted similarly across various racial-ethnic subgroups psychologists can conclude that the test assesses the same abilities regardless of a child's racial-ethnic background. If WISC-III constructs differ across racial-ethnic subgroups psychologists may not apply them the same way as they do to members of the population from which the constructs were derived.

Not only is test comparability an issue, but also fairness in testing which has sociopolitical implications (Schmeiser, 1992). Appropriate test use must examine many factors, including cultural variables, that may be related to ability expression (Geisinger, 1994). This section of the paper addresses cross-cultural conceptualizations of intelligence and the impact cultural variables may have on the expression of intelligence. Much cross-cultural research has focused on between-group test level differences. However, the focus of this paper is on qualitative differences in test performance. Although this study is not examining children from distinctly different cultures per se, these children comprising the three major racial-ethnic subgroups are from families with different cultural histories.

Sternberg (1990) discusses four models by which cross-cultural intelligence can be conceptualized and assessed. In the first model, the same intelligence test and factors underlying performance are invariant across cultures. Intelligence is expressed uniformly and observed cross-cultural differences in performance are quantitative rather than qualitative. Sternberg (1990) does allow for differential emphasis on certain aspects of intelligence across cultures. This model reflects an absolutist position where the role of culture in intellectual expression is limited and differences between groups are due to non-cultural factors (Berry, Poortinga, Segall, & Dasen, 1992).

The second model states that the same test instrument may be used to assess intelligence across cultures, but cultures differ on factors underlying test performance. The same skills and abilities are being tested, but correlational patterns between abilities vary with culture. If data obtained from different cultures are submitted to factor analysis the different factor structures that would emerge are a function of culture. Factors may differ in number or in quality. This makes unbiased qualitative comparisons of factor scores difficult (Sternberg, 1990) and misleading. This models the universalist position in cross-cultural research where the role of culture is substantial (Berry et al., 1992). The present study assessed this general conceptual model.

The third model in Sternberg's (1990) paper states that instruments used to assess intelligence differ across cultures (emic measures) but the structure of intelligence is invariant (etic constructs). Berry et al. (1992) provide a concise explanation of how etic constructs can be measured in culturally appropriate ways. This model was not directly tested in the present study. If WISC-III factor models are not generalizable across racial-ethnic subgroups, emic viewpoints on the assessment of intelligence may be considered for a more valid cross-cultural conceptualization of intelligence (i.e., the formulation of a derived etic construct; Berry et al., 1992).

The final model proposed by Sternberg (1990) is one where both the assessment instrument and factor structure underlying performance differ across cultures. This position necessitates the construction of separate tests for children of different cultures. Contextualism, or radical cultural relativism, is exemplified by this model. Given that the WISC-III is administered to children within one pluralistic society, it is not likely that this model is tenable for our purposes. Although environmental contexts do differ within this society, it is probably not to the extent that separate ability tests must be constructed for children of different racial-ethnic subgroups. In addition, children from different racial-ethnic subgroups interact which provides opportunities for integration.

Factors Influencing Cognitive Skills Development In Children From Different Racial-Ethnic Groups

Sociocultural factors often influence how and what children learn. Learning styles are shaped differently in different cultures. Nonwhite children often differ from White counterparts in their educational experiences. Hence, one would expect children from different cultural environments to express their intelligence differently. Although specific mechanisms have yet to be delineated, many cultural variables are thought to influence learning styles.

Differences in language backgrounds may not be enough to account for level differences. Spanish speaking groups perform lower academically than whites but Asian groups do well. This is probably due to complex sociocultural variables (Hakuta & Garcia, 1989). Cultural backgrounds afford individuals with specific ways of expressing knowledge. Some backgrounds will vary in their similarity to mainstream culture (Heath, 1989).

While intelligence test scores may indicate true differences in ability one must remember that tests may be insensitive to subculture values. Lower test results may reflect a minority member's distance from majority group values (Gubb & Dozier, 1989) when the majority group's values are the basis for the test's content (Hinkle, 1994). Butler-Omololu, Doster, & Lahey (1984) demonstrated the influence of cultural content of test items on test performance. Therefore, between-group variability in performance level may be more a function of sociocultural variables (Mercer, 1988) such as SES, segregated housing, achievement values, language, social participation, or acculturation rather than a function of race-ethnicity.

Even on tests exhibiting well-standardized norming techniques, racial-ethnic minority response patterns differ from the norming sample (Grubb & Dozier, 1989). Schiele (1991) and Rychlak (1995) suggest African-American epistemology is mainly affective while Euro-American epistemology is cognitive. The latter is given priority on intelligence tests. This of course fails to tap important attributes of African-Americans. Miller-Jones (1989) suggested that the interpersonal context of the testing situation may differentially impact performance across racial-ethnic subgroups. Rodriguez (1992) concurs by stating test-taking behaviors are culturally learned. The above literature emphasizes the importance of familiarity with majority group practices.

Different cultural groups have different reinforcement schedules that shape different skill patterns (Geisinger, 1992). Since specific cultural situations shape specific behaviors, one would expect different response patterns across cultures when responding to particular test items. This would argue for differential expression of cognitive abilities across different cultural groups. Problem-solving and information organizing competencies develop in the context of culturally-based activities (Miller-Jones, 1989). Empirical work is needed to determine if different abilities develop within different cultural contexts or if the same abilities develop with different modes of expression. This is one of the main problems addressed in this study.

Wechsler's Definition of Intelligence

Wechsler defined intelligence as "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" (Wechsler, 1944, p.3). Intelligence becomes manifest in a number of nonorthogonal abilities. Wechsler cited abilities as mental products sorted into classes of operation. A person's behavior is a function of how these abilities are configured and not merely a sum of their quantity.

The work of Spearman and W.B. Alexander helped shape Wechsler's conceptualization of intelligence (Wechsler, 1944). Spearman's general ("g") and specific factors theory of intelligence was tested empirically by W.B. Alexander via factor analysis. Results from this blend of theory and empirical work are reflected in the structure of Wechsler's scales.

Spearman's "g" was considered a psychometric property. It represented a factor common to all tests of intelligence. This factor accounted for shared variance between tests of intelligence. Attempting to give this psychometric entity psychological meaning, Wechsler (1944) indicated that "g" was a measure of the mind's capacity to do intellectual work. Although Spearman stated that "g" was the most important determinant of intellectual functioning in his two-factor (i.e., general and specific factors) theory, he acknowledged the existence of specific factors which accounted for variance not shared between tests of intelligence. Wechsler believed that "g" was not necessarily the most important aspect of general intelligence.

Two major lines of evidence were cited in defending the position that factors other than "g" were important determinants of intelligent behavior (Wechsler, 1944; Matarazzo, 1972). Clinical evidence suggested motivation and other conative factors were important components of problem-solving. Empirical evidence from W.B. Alexander suggested that Spearman was correct in asserting that one common factor could explain a good deal of correlational variance between intelligence tests. However, group factors as opposed to specific factors were required to explain some of the remaining variance.

Groups of intelligence tests seemed to cluster together. The abilities assessed by these tests seemed to demonstrate functional similarities. Alexander referred to these test groups as functional unities: verbal ability was one functional unity, and practical ability was another. Alexander indicated that these different functional unities were correlated. In addition to the common factor and functional unities, factors Wechsler later labeled nonintellective factors (e.g., emotion, drive) were needed to account for the final pieces of unexplained correlational variance.

The evidence cited above solidified Wechsler's position that defining intelligence on the basis of intellectual ability alone (e.g., verbal reasoning) was a mistake (Matarazzo, 1972). Practical, affective, and conative factors had an important place (Wechsler, 1950). To approximate the nature of general intelligence, one had to include "g," group factors (functional unities), and nonintellective factors. Wechsler (1944) indicated that to assess general intelligence, one must attempt to assess all of these variables. The structure of Wechsler's intelligence tests reflects this thinking.

The Verbal Scale seemed to measure intellectual skills. Verbal reasoning and abstract thinking are two examples of skills assessed by this scale. The Performance Scale was an attempt to quantify practical ability and nonintellective factors. This scale assessed one's ability to solve problems nonverbally. In addition, speed of task completion could be an indication of motivation to complete the task. This latter assessment is not objectively quantifiable in the Wechsler tests and remains left to clinical judgment.

Selecting tests to assess these aspects of intelligence was based on three criteria (Wechsler, 1944). First, Wechsler inspected tests in use. He also checked the tests' validity and clinical utility. Wechsler acknowledged that no test battery could measure every aspect of general intelligence. However, he thought that it was possible to assess a sufficient portion to get a fairly reliable estimate of a person's functioning. The tests Wechsler selected for his first adult intelligence battery have remained largely the same through the years. The children's tests remain a downward extension of the adult versions. The section below describes the subtests of Wechsler's children's battery.

The WISC-III Wechsler, 1991) is the latest intellectual assessment device for children in the Wechsler series. Its predecessors, the WISC (Wechsler, 1949) and WISC-R (Wechsler, 1974), proved to be widely used and exhaustively researched clinical instruments. Many exploratory and confirmatory factor analytic studies assess the test's construct validity across a variety of samples (e.g., racial-ethnic groups, children referred for psychoeducational assessment). This study assesses the invariance of factor models underlying the WISC-III in White, Black, and Hispanic subgroups of the standardization sample.

Organization of the Children's Scales

The structure of the children's test reflects David Wechsler's conceptualization of intelligence as a multidimensional construct (Wechsler, 1949). Many different abilities comprise intelligence which allow for the individual to behave effectively within the environment (Wechsler, 1939). No single ability was deemed more important than any other. Wechsler therefore included many types of subscales in his tests to measure different aspects of intelligence.

Like earlier versions of the test, the WISC-III bifurcates into Verbal and Performance scales. These scales assess two major areas of cognitive ability: verbal ability and nonverbal ability (both described below). Both sets of abilities may manifest themselves in many ways. Hence, each scale contains five routinely administered subtests plus one optional verbal subtest (Digit Span) and two optional nonverbal subtests (Mazes and Symbol Search). These optional subtests do not contribute to IQ scores but yield additional clinical information about the examinee. Each subtest is purported to measure a specific ability.

Collectively, the WISC-III Verbal Scale subtests assess the application of verbal skills to problem solving situations (Sattler, 1992). Each Verbal Scale subtest purports to measure a specific verbal ability. The subtests and the abilities they assess include: (a) Vocabulary (word knowledge); (b) Similarities (logical verbal abstract reasoning); (c) Information (range of general factual knowledge); (d) Comprehension (knowledge of conventional standards of behavior); (e) Arithmetic (quantitative reasoning); and (f) Digit Span (short-term auditory sequential recall).

The Performance Scale assesses one's skill at thinking in visual images and manipulating these images fluently (Sattler, 1992). Individual subtests assess specific nonverbal abilities. The subtests and abilities assessed include: (a) Picture Completion (visual perception of essential detail); (b) Picture Arrangement (ability to perceive causal relationships); (c) Coding (efficiency on a rote visual-motor task); (d) Block Design (nonverbal concept formation); (e) Object Assembly (interpretation of relationships among parts); (f) Symbol Search (visual discrimination skills); and (g) Mazes (planning ability).

One can see that verbal and nonverbal abilities are not unitary concepts. Both types of abilities may manifest themselves in many ways. Individuals may vary in how they express these abilities. Such variation is reflected in subjects' relative performance on different subtests.

Wechsler tests generate three different I.Q. scores. The score often associated with a person's general intelligence is the Full Scale I.Q. This score results from a person's overall performance across verbal and performance subtests. A Verbal Scale I.Q. and Performance Scale I.Q. can also be derived based on a person's performance on subtests within each scale. Like the WISC and WISC-R, the WISC-III generates deviation I.Q.'s. WISC-III deviation I.Q.'s result by comparing an examinee's performance to the performance of a representative sample of children aged 6-16 which closely resembles the 1988 United States' census data (Wechsler, 1991).

The organization of the WISC-III resembles that of the WISC and WISC-R. Most modifications of successive editions were of item content, the addition of new supplemental subtests, and updating norms. The tests maintain the verbal-performance dichotomy. As the test was modified over the years, factor analysis determined if the WISC tests retained their construct validity.

Factor Analysis and Wechsler Tests

Factor analysis has long been the preferred method for assessing the construct validity of clinical instruments. The method seeks to determine latent structures that may account for observed interrelationships amongst variables. In the case of intelligence tests, the investigator attempts to determine the nature of cognitive abilities responsible for test performance.

Exploratory factor analyses of the WISC, WISC-R, and WISC-III pervade the empirical literature. Different methods of exploratory factor extraction exist such as principal components and common factor analysis. Different rotation techniques (e.g. orthogonal versus oblique) exist as well. Researchers often make arbitrary atheoretical decisions as to which method of factor analysis to use. Criteria for determining the significance of a factor are also arbitrary.

Sound rationales for using various techniques are sometimes offered. However, using different techniques across studies can produce inconsistent findings (Floyd & Widaman, 1995). Such inconsistencies may include differences in number of factors retained and differences in factor loading patterns. This makes cross-study comparisons difficult. One may wonder if different results across studies are a function of the sample studied or the factoring method itself.

Recently, confirmatory factor analysis has been used to assess construct validity (e.g., Joreskog & Sorbom, 1989; Pedhazur & Schmelkin, 1991). With confirmatory analysis, models are proposed a priori that are said to account for test performance. This method avoids the problems cited above but has other limitations (e.g., see Floyd & Widaman, 1995). Confirmatory factor analysis has been applied to WISC-R and WISC-III research.

The critical review below integrates the findings mentioned in cited works with criteria applied by the present author. To maintain consistency across studies, the present author considers a factor loading of .30 significant. This is a commonly accepted value in the psychometric literature. Frequently, individual studies of the WISC, WISC-R, and WISC-III reveal subtests with secondary factor loadings. For example, Subtest A may load .30 or greater on both Factor X and Factor Y. While this splitting of subtest variance between factors is not discussed within empirical articles it is taken into consideration in this paper.

This splitting of significant variance between two (or more) factors is theoretically and practically important. Theoretically, these results challenge Wechsler's conceptualization of intelligence as reflected in the organization of his test. These results also have implications for the validity of a child's clinical profile based on Wechsler tests.

To illustrate, suppose Subtest A loads significantly only on Factor X in one study and only on Factor Y in one study. In seven studies, it loaded significantly on both Factors X and Y. However, in four of the seven, Subtest A had higher factor loadings on Factor X. For purposes of this review, Subtest A would be said to load on Factor X since, across nine studies, more of its variance seems to be determined by Factor X. This does not eliminate the problem of splitting significant variance. However, this is an effort to maintain parsimony in interpreting factors underlying Wechsler tests across many samples.

WISC Factor Structure: Standardization Data

The WISC was standardized on 100 boys and 100 girls at each age from 5 to 15 years inclusive (Wechsler, 1949). The sample was stratified across geographic area, urban-rural residence, and parental occupation. Race-ethnicity was not a stratification variable. The final sample of 2200 children included White children only. The sample was roughly representative of the 1940 United States' census across these variables.

The original WISC manual did not report data regarding the test's factor structure. Wechsler (1949) provided evidence of acceptable levels of reliability for the subtests, Full Scale, Verbal Scale, and Performance Scale scores. Cohen (1959), following his work on the Wechsler adult tests, seemed to initiate the first in a long line of factor analytic studies of the Wechsler children's tests.

Data analyzed by Cohen (1959) were those reported in the Wechsler (1949) subtest intercorrelation tables. Although the test was standardized on children aged 5-15 years, these tables provided subtest scaled score intercorrelations only for children aged 7.5, 10.5, and 13.5 years. Data were factor analyzed by the complete centroid method and rotated obliquely. Factor loadings of .20 or greater were deemed sufficient for factor interpretation

(Cohen, 1959)

Cohen (1959) extracted and labeled five first-order factors and one second-order factor. Since their content and nomenclature resemble contemporary factors, they merit some discussion. Labels assigned to these factors originated in research of early Wechsler adult scales (e.g., Wechsler Adult Intelligence Scale; Wechsler, 1955): Verbal Comprehension I, Perceptual Organization, Freedom From Distractibility, Verbal Comprehension II, and "Factor E" (quasi-specific). Second-order factor analysis indicated that all subtests loaded substantially on a "g" or general intelligence factor (Cohen, 1959).

Information and Similarities loaded on Verbal Comprehension I at each age level. Arithmetic (7.5 and 10.5) and Vocabulary (7.5 and 13.5) each loaded at two age levels. Comprehension loaded on this factor at the 13.5 age level. Cohen (1959) deemed this factor barely distinguishable from Verbal Comprehension II. However, Verbal Comprehension I suggests verbal knowledge gained through formal education (Cohen, 1959). Factor loadings resembled those on the Verbal Comprehension factor of the WAIS, especially at age 13.5.

The second factor extracted was Perceptual Organization. Block Design and Object Assembly loaded significantly on this factor at each age level. Picture Completion (10.5 and 13.5) and Mazes (7.5 and 10.5) loaded at two age levels. Picture Arrangement loaded on this factor in the 7.5 year age level. This factor also resembled its <u>WAIS</u> counterpart (Cohen, 1959).

Digit Span loaded at each age level on the third factor, Freedom From Distractibility. Mazes loaded at the 10.5 and 13.5 age levels (but less than .30). Arithmetic (13.5), Picture Arrangement (7.5), and Object Assembly (10.5) loaded at one age level. Cohen (1959) indicated that Arithmetic and Digit Span comprised the adult Freedom From Distractibility Factor for the <u>WAIS</u>. However, for the <u>WISC</u>, only Digit Span and Mazes (as per Cohen, 1959) load consistently on this factor in at least two age levels.

Comprehension and Picture Completion (as per Cohen even though the subtest loaded less than .30) loaded significantly on the fourth factor, Verbal Comprehension II at each age level. Vocabulary loaded at the 7.5 and 10.5 age levels. Similarities loaded at the 13.5 age level. Differentiating it from the other verbal factor, Cohen (1959) suggested that this factor reflects an application of verbal skills to new situations.

The fifth factor to emerge was labeled Quasi-specific since no unifying theme was ascertainable. Coding loaded on this factor at each age level. Picture Arrangement loaded on the Quasi-specific factor at ages 10.5 and 13.5. The relevance of, and existence of, this factor faded over the years.

A second-order factor analysis revealed that a general factor accounted for the first order factor intercorrelations. Cohen (1959) indicated satisfactory subtest correlations with this general or "g" factor (median correlation of .58 across age levels). Thus, the Full Scale I.Q., a combination of subtest scaled scores, appeared to be a good estimate of "g" or general intelligence.

One must interpret this work with caution. The WISC standardization sample was stratified only on geographic location, urban-rural residence, and parental occupation. While the sample closely resembled 1940 United States' census data (Wechsler, 1949), the sample included White children only. Results of the above factor analysis cannot be generalized to children of different racial-ethnic groups.

Silverstein (1969) argued that Cohen's (1959) five-factor solution lacked descriptive efficiency. Silverstein (1969) also examined WISC standardization data with a different method. Principal factor analysis with oblique Maxplane rotation of two factors suggested that two factors may explain WISC subtest intercorrelations. This study suggests that factoring methods may partially determine the number of factors retained.

Silverstein's (1969) factors represented combinations of Cohen's (1959) factors. The verbal factor loaded all Verbal Scale subtests. This factor seemed to blend Cohen's

(1959) Verbal Comprehension I and Verbal Comprehension II factors. All Performance Scale subtests, except Picture Arrangement and Coding, loaded on the performance factor. Picture Arrangement loaded on the verbal factor and Coding did not load significantly on either factor. Silverstein's performance factor represented a blend of Cohen's (1959) Perceptual Organization and Quasi-specific factors. Silverstein (1969) concluded that the two-factor solution exhibited statistical invariance across Wechsler tests.

Blaha, Wallbrown, and Wherry (1974) examined WISC factor structure with a hierarchical factoring technique advocated by Wherry (1959). Analysis of the standardization data yielded support for a higher order "g" factor and two lower order verbal and performance factors.

The verbal factor loaded all Verbal Scale subtests. However, the Digit Span subtest exhibited a loading less than .30. The performance factor loaded all Performance Scale subtests. Picture Completion, Picture Arrangement, and Coding exhibited loadings less than .30. Evidence for a Freedom From Distractibility factor was not investigated as it did not fit the author's theoretical orientation.

Summary of WISC Factor Analytic Research: Standardization Data

In sum, investigations of the WISC standardization data yielded strong evidence for general, verbal, and performance factors. Table 1 shows results of the three studies described above. In the table, subtests are assigned to factors if they exhibited factor loadings exceeding .30. Across studies, the verbal factor was composed of Information, Comprehension, Arithmetic, Similarities, and Vocabulary. The performance factor was consistently comprised of Block Design, Object Assembly, and Mazes. Picture Completion joined this factor in Cohen (1959) and Silverstein (1969). Digit Span, Picture Arrangement, and Coding did not load consistently on any particular factor. Evidence for the existence and composition of the Freedom From Distractibility factor was weak in Cohen (1959).

Wechsler's partitioning of intelligence into verbal and nonverbal factors was generally supported. However, evidence for nonintellective factors such as Freedom From Distractibility, was weak. In addition, three subtests noted above did not load consistently on a particular factor. Thus empirical data above provided mixed support for Wechsler's theory.

WISC Factor Structure in Different Racial-Ethnic Groups

Semler and Iscoe (1966) investigated the factor structure of the WISC in samples of White (N=141) and Black (N=134) public school children aged 5-9 years. Subtest intercorrelations (excluding Comprehension, Picture Arrangement, and Mazes) were subjected to an exploratory maximum likelihood factor analysis with varimax rotation. Results suggested a three-factor solution for both groups.

Table 1

WISC Standardized Subtest Factor Loadings: Standardization Data

	1*		2		3	
Subtest	Factor	Loading[1]	Factor	Loading	Factor	Loading
I	VI	.32-.41	VC	.67	VC	.42
C	VII	.20-.47	VC	.57	VC	.34
A	VI	.07-.41	VC	.57	VC	.34
S	VI	.26-.38	VC	.60	VC	.36
V	VI&II	.21-.47/-.01-.30	VC	.65	VC	.39
DSp	FFD	.33-.44	VC	.41	None	NS
PC	PO	.06-.50	PO	.33	None	NS
PA	Q	-.10-.38	VC	.31	PO	.33
BD	PO	.46-.55	PO	.53	PO	.45
OA	PO	.49-.59	PO	.62	PO	.54
CD	Q	.23-.39	None	NS	None	NS
MZ	PO	.42-.56	PO	.40	PO	.33

Note. *1=Cohen (1954); 2=Silverstein (1969); 3=Blaha et al. (1974).

[1] Range of loadings ages 7.5, 10.5, and 13.5.

I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; VI=Verbal Comprehension I; VII=Verbal Comprehension II; VC=Verbal Comprehension; PO=Perceptual Organization; FFD=Freedom From Distractibility; Q=Quasi-specific; None=no significant loading; NS= factor loading < .30.

Factor loadings differed substantially across groups on two of the three factors. Additionally, the factor structures obtained by Semler and Iscoe (1966) differed substantially from structures expected from Wechsler's theory.

For both racial-ethnic subgroups the verbal factor was comprised of Information, Similarities, Vocabulary, and Picture Completion. Digit Span also loaded on this factor for the Black group. For the White group Arithmetic, Digit Span, Block Design, and Coding clustered together as a blend of performance and Freedom From Distractibility factors. In the Black group, Block Design, Object Assembly, and Coding clustered as a performance factor. Arithmetic stood alone as a factor for the Black group. Object Assembly stood alone for the White group.

Silverstein (1973) assessed <u>WISC</u> factor structure across samples of 6-11 year-old Anglo (<u>N</u>=505), Black (<u>N</u>=318) and Chicano (<u>N</u>=487) public school children. Each group's subtest correlation matrix was factored with ones in the diagonal. Results suggested a two-factor solution for each group. Reanalysis by principal factoring with oblique maxplane rotation generated similar factors across groups. Silverstein (1973) cited a median coefficient of congruence of .95. However, factor loadings warrant closer inspection as only the verbal factor seemed to conform generally to Wechsler theory across groups.

The first factor appeared to be a verbal factor.

Information, Comprehension, Arithmetic, Similarities, and Vocabulary loaded on this factor for each group. Digit Span loaded on this factor for the Black sample. Digit Span did not load on either factor for the Anglo and Chicano groups.

The second factor appeared to be a performance factor. Block Design and Object Assembly loaded on this factor for each group. However, Picture Completion and Picture Arrangement also loaded on this factor for the Black group. These subtests failed to load significantly on either factor for the Anglo and Chicano groups. Coding did not have a significant factor loading across racial-ethnic groups.

A hierarchical factor analysis by Vance, Huelsman, & Wherry (1976) was applied to data generated by 10-11 year-old disadvantaged White (N=60) and Black (N=30) children. Intercorrelations of 10 WISC subtests (excluding Digit Span and Mazes) were factor analyzed. Results indicated the absence of a general factor possibly due to a restricted range of Full Scale I.Q. (Vance et al., 1976). The two largest factors resembled verbal and performance factors. For each subgroup, all performance subtests loaded highest on the performance factor. All verbal subtests, except for Comprehension, loaded highest on the verbal factor. One must interpret these results with caution because of small sample sizes.

Summary of WISC Factor Analytic Research: Racial-Ethnic Subgroup Data

Tables 2, 3, and 4 show summaries of the above research on White, Black, and Hispanic subgroup data. In sum, Wechsler's theoretical verbal-performance bifurcation was generally not well supported. The verbal factor, especially in Silverstein's (1973) oblique two-factor solution, seemed to fare better than the performance factor across studies. In Semler & Iscoe's (1966) orthogonal three-factor solution, both verbal and performance factors occasionally loaded the other's subtests across White and Black samples. Differences in factor structure across racial-ethnic groups exist regardless of rotation method. It is possible that noninclusion of subtests from analysis and the study of smaller sample sizes produced the discrepancies between factor analytic studies of the standardization sample and studies of racial-ethnic subgroups. However, real differences in structure of intelligence between racial-ethnic subgroups cannot be ruled out.

Table 2 shows studies from White subgroups. When both studies are considered (Vance et al., 1976 was excluded because of small sample size) the verbal factor consisted of Information, Comprehension, Similarities, and Vocabulary. The performance factor consisted of Block Design and Object Assembly. Arithmetic, Digit Span, Picture Completion, Picture Arrangement, and Coding did not load consistently on

Table 2

WISC Standardized Subtest Factor Loadings From White Subgroup Data

	1[*]		2[1]	
Subtest	Factor	Loading	Factor	Loading
I	VC	.70	VC	.61
C	--	--	VC	.51
A	PO/FFD	.62	VC	.44
S	VC	.62	VC	.56
V	VC	.73	VC	.64
DSp	PO/FFD	.46	None	NS
PC	VC	.43	None	NS
PA	---	--	None	NS
BD	PO/FFD	.65	PO	.38
OA	PO	.98	PO	.46
CD	PO/FFD	.54	None	NS
MZ	--	--	--	--

Note. [*] 1=Semler & Iscoe (1966); 2=Silverstein (1973). [1]Two-Factor solution specified. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; VC=Verbal Comprehension; PO=Perceptual Organization; FFD=Freedom From Distractibility; --= subtest not examined; None=did not load significantly on a factor; NS=factor loading less than .30.

any factor. Mazes was not studied.

Table 3 shows studies from Black subgroups. As Table 3 indicates, the verbal factor consisted of Information, Comprehension, Similarities, Vocabulary, and Digit Span. The performance factor consisted of Picture Arrangement, Block Design, and Object Assembly. Arithmetic, Picture Completion, and Coding did not load consistently on any factor. Mazes was not studied.

Table 4 shows results from the only study done on an Hispanic subgroup. The verbal factor consisted of Information, Comprehension, Arithmetic, Similarities, and Vocabulary. The performance factor consisted of Block Design and Object Assembly. Digit Span, Picture Completion, Picture Arrangement, and Coding did not load significantly on any factor. Again, Mazes was not studied.

WISC-R Factor Structure: Unrestricted Analyses of Standardization Data

Unlike the WISC, the WISC-R (Wechsler, 1974) standardization sample included children of various racial-ethnic groups. A child's racial-ethnic status was characterized as either White or Nonwhite. Nonwhite members included children from Black, American Indian, and Oriental groups. Puerto Rican and Chicano children were categorized as White or Nonwhite based on a vague "visible physical characteristics" criterion (Wechsler, 1974).

The final standardization sample consisted of 2200

Table 3

WISC Standardized Subtest Factor Loadings From Black Subgroup Data

	1*		2[1]	
Subtest	Factor	Loading	Factor	Loading
I	VC	.54	VC	.54
C	--	--	VC	.48
A	FFD	.98	VC	.49
S	VC	.77	VC	.47
V	VC	.71	VC	.56
DSp	VC	.39	VC	.37
PC	VC	.57	PO	.33
PA	---	--	PO	.34
BD	PO	.57	PO	.49
OA	PO	.97	PO	.53
CD	PO	.31	None	NS
MZ	--	--	--	--

Note. * 1=Semler & Iscoe (1966); 2=Silverstein (1973). [1]Two-Factor solution specified. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; VC=Verbal Comprehension; PO=Perceptual Organization; FFD=Freedom From Distractibility.
--= subtest not examined; None=did not load significantly on a factor; NS=factor loading less than .30.

Table 4

WISC Standardized Subtest Factor Loadings From Hispanic Subgroup Data

	1[a]	
Subtest	Factor[1]	Loading
I	VC	.58
C	VC	.38
A	VC	.39
S	VC	.54
V	VC	.60
DSp	None	NS
PC	None	NS
PA	None	NS
BD	PO	.45
OA	PO	.45
CD	None	NS
MZ	--	--

Note. [a]1=Silverstein (1973); [1]Two-factor solution specified. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; VC=Verbal Comprehension; PO=Perceptual Organization; --= subtest not examined; None=did not load significantly on a factor; NS=factor loading less than .30.

children (100 male and 100 female at each of 11 age levels) and approximated the 1970 United States' census data. Stratification variables included age, sex, race-ethnicity, geographic region, urban-rural residence, and head of household occupation (Wechsler, 1974).

Subtest intercorrelations for each of the 11 age levels are presented in the WISC-R manual. Evidence attests to adequate levels of subtest reliability (Wechsler, 1974). These data provided the base for numerous empirical studies of the WISC-R factor structure.

Kaufman (1975) performed exploratory factor analyses of the WISC-R standardization data. Factor analyses were performed on each of the eleven age levels of the standardization sample. This study employed different factoring and rotation techniques in order to: (a) have an objective "guide" to determine the number of factors to rotate; and (b) to rule out the possibility that factors obtained were a function of specific rotation procedures. First, a principle components factor analysis (ones on the diagonal) was performed with varimax rotation of factors (Kaiser-Guttman criterion). This procedure generated significant Verbal and Performance factors at six age levels. It produced Verbal Comprehension, Perceptual Organization, and Freedom From Distractibility factors at five age levels.

The study's second phase used principal factor analysis (squared multiple correlations on the diagonal) at each age level followed by varimax, oblimax, and biquartimin rotation of two-, three-, four-, and five- factor solutions. The two-factor rotated solution resulted in Verbal and Performance factors at all age levels (Kaufman, 1979). They were named Verbal Comprehension and Perceptual Organization to maintain consistency with previous work (e.g., Cohen, 1959). These factors closely corresponded to the Verbal-Performance dichotomy of the WISC-R scale.

The three-factor rotated solution produced Verbal Comprehension and Perceptual Organization factors. In addition, a third factor emerged in 9 of 11 age levels, Freedom From Distractibility. The four-factor rotated solution produced the above three factors at all 11 age levels plus a Quasi-specific factor at six age levels. The five-factor solution added little more to the above results.

Kaufman (1975) concluded that Verbal Comprehension, Perceptual Organization, and Freedom From Distractibility are meaningful factors underlying the WISC-R. The number and composition of factors that emerged throughout the study were generally consistent. They did not seem to be a function of factoring method: orthogonal and oblique rotations produced similar results).

Within the Verbal Comprehension factor, Vocabulary, Information, Comprehension, and Similarities had the highest loadings across the age range. Arithmetic was the fifth best measure of this factor. Verbal Comprehension closely resembled the WISC-R Verbal Scale. The WISC-R Verbal Comprehension factor appears more stable across the age range than did its WISC counterpart which split into Verbal Comprehension I (Information, Comprehension, Similarities, and Vocabulary) and Verbal Comprehension II (Comprehension, Picture Completion, Vocabulary at ages 7.5 and 10.5, and Similarities at age 13.5).

Block Design, Object Assembly, and Picture Completion emerged as the best measures of the Perceptual Organization factor. All Performance Scale subtests with the exception of Coding had median factor loadings of .40 or greater across the age range. Perceptual Organization closely resembled the WISC-R Performance Scale. This factor was also similar to the WISC Perceptual Organization factor.

The Freedom From Distractibility factor was composed of Arithmetic and Digit Span at each age level. Information and Coding also loaded on this factor at most age levels. Given these results, Kaufman (1975) indicated that this factor was difficult to interpret but still meaningful. Kaufman's final Freedom From Distractibility factor (Arithmetic, Digit Span, and Coding) was somewhat similar to Cohen's (1959); Digit Span, Mazes (at 10.5 and 13.5 years), and Arithmetic, Picture Arrangement, and Object Assembly at one age level each.

Kaufman (1975) cited evidence supporting the Full Scale

I.Q. His first unrotated principal factor accounted for a median 82% of common factor variance. Nine of the 12 subtests loaded .60 or better on this general factor. Vocabulary (.80) was the best measure of this factor while Coding (.41) was the worst. In addition, Kaufman indicated that oblique factors were significantly correlated which may be explained by a common higher order factor. Like Cohen's (1959) WISC study, Kaufman (1975) found evidence supporting the WISC-R Full Scale I.Q. as a good measure of general intelligence.

Replications of Kaufman's (1975) study appear in the literature. A hierarchical factor analysis of WISC-R standardization data indicated a general factor emerges before two subgeneral verbal and performance factors (Wallbrown, Blaha, Wallbrown, & Engin, 1975). Silverstein's (1977) principal factor analysis suggested a negligible difference between oblique two- (Verbal and Performance) and three-factor (Verbal, Performance, and Freedom From Distractibility) solutions.

Summary of WISC-R Factor Structure: Unrestricted Analyses of Standardization Data

Table 5 shows results from factor analyses of WISC-R standardization data. Two-factor solutions by Wallbrown et al. (1975) and Silverstein (1977) noted some consistency despite different factoring techniques. Both studies noted the verbal factor to contain Information, Comprehension,

Table 5

WISC-R Subtest Factor Loadings: Standardization Data

	1a	b	c*	a	b	c	2[1]		3a	b	a	b
Subtest	Factor			Loading			Factor	Loading	Factor		Loading	
I	V/F	V	V	.63/.41	.69	.63	V	.35	V	V	.53	.42
S	V/PO	V	V	.64/.34	.66	.62	V	.30	V	V	.49	.40
A	V/F	F	F	.37/.58	.48	.43	V	.30	V	F	.47	.30
V	V/F	V	V	.72/.33	.81	.71	V	.41	V	V	.64	.47
C	V/PO	V	V	.64/.30	.69	.64	V	.30	V	V	.45	.45
DSp	F	F	F	.56	.55	.56	None	NS	V	F	.40	.43
PC	V/PO	PO	PO	.35/.57	.54	.55	PO	.31	PO	PO	.45	.38
PA	V/PO	PO	PO	.33/.41	.45	.39	None	NS	PO	PO	.33	.30
BD	PO	PO	PO	.66	.68	.66	PO	.40	PO	PO	.51	.46
OA	PO	PO	PO	.65	.69	.67	PO	.40	PO	PO	.55	.53
CD	F	F	F	.42	.38	.39	None	NS	None	F	NS	.30
MZ	PO	PO	PO	.47	.53	.48	None	NS	PO	PO	.39	.36

Note. *1=Kaufman (1975), a=Varimax rotation, b=Oblimax, c=Biquartimin; 2=Wallbrown et al. (1975); 3=Silverstein (1977), a=Two-factor solution, b=Three-factor solution. [1]Two-factor solution only. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; V=Verbal Comprehension; PO=Perceptual Organization; F=Freedom From Distractibility; None= did not load on a factor; NS= factor loading < .30.

Arithmetic, Similarities, and Vocabulary. The performance factor consisted of Picture Completion, Block Design, and Object Assembly. No consistency was noted for Digit Span, Picture Arrangement, Coding, and Mazes.

As Table 5 shows, the oblique (Kaufman, 1975; Silverstein, 1977) and orthogonal (Kaufman, 1975) three-factor solutions of the WISC-R remained consistent across studies. Information, Similarities, Vocabulary, and Comprehension consistently comprise the Verbal Comprehension factor. Block Design, Object Assembly, Picture Completion, Picture Arrangement, and Mazes make up the Perceptual Organization factor. Arithmetic, Digit Span, and Coding comprise the Freedom From Distractibility factor. Studies compiled by Kaufman (1979b) also cite stability for the Freedom From Distractibility factor.

Also seen in Table 5 is that Kaufman's (1975) orthogonal solution produced many instances of secondary factor loadings. The oblique three-factor solutions reviewed did not produce any secondary loadings. Hence, oblique solutions may be more appropriate when analyzing WISC-R factor structure. Restricted factor analyses discussed below concur with this analysis.

WISC-R Factor Structure: Restricted Analyses of Standardization Data

O'Grady (1989) used a simultaneous maximum-likelihood confirmatory factor analysis to determine if the WISC-R factor structure could be explained by a one-, two-, or three-factor solution. Intercorrelations between the 12 WISC-R subtests were analyzed for all 11 age groups of the standardization sample. Replication followed on 9 published data sets involving factor analysis of the WISC-R.

O'Grady (1989) tested models described in the literature. The one-factor model contained loadings for all 12 subtests. Orthogonal and oblique multifactor models were also tested: a two-factor verbal and performance model and a three-factor Verbal Comprehension, Perceptual Organization, and Freedom From Distractibility model. Evaluation and comparison of models were done with chi-square goodness of fit tests, a Goodness of Fit Index, Bentler-Bonnet's delta, and a chi-square difference test.

All models except orthogonal two- and three-factor models, which demonstrated poorer fit, did an equally good job of fitting the data across age groups. Oblique two- and three-factor models demonstrated an improvement in fit over the single factor model. The oblique three-factor demonstrated improved fit over the two-factor model. However, these improvements in fit were minimal (O'Grady, 1989) and perhaps inconsequential. For multifactor models, unique variances were not consistent across age groups. Factor variances and covariances were consistent across the age groups.

The single factor and oblique multi-factor models were tested in the replication. Goodness of fit tests indicated that none of the models fit the data well. The single factor model fit better than the null model. The oblique two-factor model showed better fit than the single factor and null models. Finally, the oblique three-factor model showed an improved fit over the single factor and null models. The oblique three-factor model fit better than the oblique two-factor model only in three of the seven studies that produced interpretable results (O'Grady, 1989).

The author cited a preference for a one-factor solution. While the oblique multi-factor models provided a better fit, they did so minimally. O'Grady (1989) indicated that any model with relaxed constraints would do so. The author accepted a parsimonious one-factor solution.

Additionally, the oblique models in this study evidenced some degree of misspecification. O'Grady concluded that this weakens both the argument for the <u>WISC-R</u> verbal-performance dichotomy and evidence for a Freedom From Distractibility factor. Given the poor fit of orthogonal models, the study also questions the rationale for orthogonal rotations used in many studies.

Macmann & Barnett (1992) used both exploratory and confirmatory factor analyses to examine <u>WISC-R</u> standardization data. This study, like O'Grady's (1989), defends a one-factor solution. This conclusion was based primarily upon the facts that: (a) multifactor models produced little incremental fit over a one factor solution; (b) the verbal and performance factors were significantly correlated; and c) subtests loaded highly on both verbal and performance factors.

Summary of WISC-R Factor Analytic Research:Standardization Data

As a whole, the above research of the WISC-R standardization data tend to support Wechsler's conceptualization of intelligence. The WISC-R seemed to be a better fit to Wechsler's theory of intelligence than its predecessor. Exploratory factor analytic studies of the WISC-R standardization data provide evidence for Verbal Comprehension, Perceptual Organization, and Freedom From Distractibility factors. Restricted (confirmatory) factor analyses, more direct tests of theoretical models, indicated that two- and three-factor solutions seem to fit the data better than the one-factor solution, but only marginally. Authors of the restricted analyses preferred more parsimonious one-factor solutions.

Acceptance of a one-factor solution despite improved fit of multifactor solutions not only contradicts Wechsler's theory but affects clinical utility of tests. Unless multifactor models can be outright rejected because of poor model fit, it may be wise to retain multifactor models. If the validity of a two- or three-factor model is reasonably substantiated (as noted above) and accepted (not necessarily noted above) more precise clinical statements may be made about a child's functioning. For example, instead of saying that a child's overall functioning is low (one-factor model) we may say, a child's overall functioning is low but her verbal ability is much better developed than her nonverbal ability (two-factor model). While parsimony is certainly a criterion to be included in decisions involving scientific acceptance of theoretical models, it is by no means the only one nor is it the most important.

WISC-R Factor Structure in Racial-Ethnic Subgroups

A number of studies investigated the generalizability of the one-, two-, and three-factor models in ethnically diverse samples. This section presents the authors' findings from exploratory factor analyses which used orthogonal rotations only. In the summary that follows, general trends will be discussed.

Reschly (1978) first studied the WISC-R factor structure across diverse racial-ethnic samples. Data generated by Anglo (*N*= 252), Black (*N*= 235), Chicano (Mexican-American; *N*= 223), and Native American Papago children (*N*= 240) from Arizona were analyzed. Different exploratory factor analytic techniques consistent with previous literature (e.g., Kaufman, 1975) were used.

A principal components analysis was performed separately for each racial-ethnic group. The Kaiser-Guttman criterion determined the appropriate number of factors.

Next, an unrestricted maximum likelihood factor analysis was performed for two-, three-, and four-factor solutions. At each step, a chi-square goodness of fit test compared obtained factor matrices to each group's WISC-R subtest correlation matrix.

Principal factor analysis was performed separately for each group. Varimax rotation was used for two-, three-, and four- factor solutions regardless of the appropriate number of factors to rotate. This procedure allowed the study to correspond to previous work which rotated as many as four factors (Reschly, 1978).

Principal components analysis suggested a three-factor solution for Anglo and Chicano children. A two-factor solution emerged for Black and Native American Papago children. Goodness of fit tests confirmed that more than two factors were needed to account for the Anglo subtest correlation matrix. However, only two factors were required to explain Chicano, Black, and Native American Papago correlation matrices. Reschly (1978) provided data only for the three-factor solution across subgroups.

The two-factor solution indicated verbal and performance factors. Vocabulary, Information, Comprehension, and Similarities loaded highest on the verbal factor for each group. Additionally, all verbal subtests loaded significantly on this factor. Block Design and Object Assembly loaded highest on the performance factor across groups. All performance subtests loaded significantly on this factor with the exception of Coding. Coefficients of congruence ranged from .97 to .99. This indicated similarity of factors across racial-ethnic groups.

When the three-factor solution was analyzed, different results emerged. Secondary loadings were noted for many subtests within each racial-ethnic group. The three-factor solution for the Anglo group resembled the solution obtained by Kaufman (1975). Coefficients of congruence ranged from .97 to .98.

For Chicanos, the Verbal Comprehension factor consisted of Information, Similarities, Vocabulary, Comprehension, and Digit Span. The Perceptual Organization factor consisted of Picture Completion, Block Design, Object Assembly, and Mazes. The Freedom From Distractibility factor consisted of Arithmetic, Picture Arrangement, and Coding which did not resemble Kaufman (1975).

For the Black group, the Verbal Comprehension factor consisted of Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span. The Perceptual Organization factor consisted of Picture Completion, Picture Arrangement, and Mazes. Object Assembly and Block Design stood alone as another factor. This solution does not resemble Kaufman's (1975) three-factor solution. Native American Papago data generated Verbal and Performance factors similar to Kaufman's (1975), but the third factor consisted of only the Mazes subtest.

Reschly (1978) indicated that the first unrotated principal first factor accounted for 79%, 83%, 79%, and 77% of the variance for Anglo, Black, Chicano, and Native American Papago groups, respectively. An unrotated first principal component yielded similar results. Reschly (1978) also performed a restricted maximum likelihood procedure which partitioned unique variance associated with verbal and performance factors from the general factor. He indicated that variance accounted for across Anglo, Black, Chicano, and Native American Papago groups was 61%, 63%, 59%, and 61%, respectively.

The study's support for the general factor across racial-ethnic samples was similar to support found for a general factor in the WISC-R standardization sample. However, Reschly (1978) concluded that the factor structure of the WISC-R may be different across racial-ethnic groups. Most notably, the study failed to support the Freedom From Distractibility factor in the Black and Native American Papago samples. Support for the Full Scale I.Q. and a verbal-performance dichotomy of the WISC-R were found.

Vance and Wallbrown (1978) investigated WISC-R factor structure in a sample of 150 Black children aged 6-15 years. Intercorrelations of ten subtests (excluding Digit Span and Mazes) were subjected to a hierarchical factor analysis. Two factors were specified to control factorization.

Results supported a weak higher order "g" factor. Verbal and performance factors (not involving Coding) were also derived.

Stedman, Lawlis, Cortner, & Achterberg (1978) studied a referred sample of which 90% had Hispanic surnames. Principal components analysis of 11 subtests (excluding Mazes) with varimax rotation yielded factors similar to Kaufman (1975). Again interpretation warrants caution because of small sample size (N = 106).

A relatively large unrotated first principal component emerged which provided evidence for a general factor. The Verbal Comprehension factor consisted of Information, Similarities, Arithmetic, Vocabulary, and Comprehension. The Perceptual Organization factor consisted of Picture Completion, Picture Arrangement (whose highest loading was actually on the verbal component), Block Design, and Object Assembly. The Freedom From Distractibility factor consisted of Arithmetic, Digit Span, and Coding. Significant secondary loadings were found for Arithmetic, Vocabulary, Digit Span, Picture Arrangement, and Object Assembly.

Gutkin and Reynolds (1980) investigated the factor structure of the WISC-R by examining data from Anglo (N= 78) and Chicano children (N= 142) in the southwest United States. Principal factor analysis was performed separately for each group. Intercorrelations of standard scores from 11 WISC-R subtests were factored (excluding Mazes). Factors exceeding an eigenvalue of 1.0 were rotated by the varimax method.

Results indicated a two-factor solution for each group. Coefficients of congruence were .98 and .91. A first unrotated principal factor exhibited a .99 coefficient of congruence. For both groups, the first rotated factor, Verbal Comprehension, contained Information, Similarities, Vocabulary, and Comprehension. Arithmetic loaded highest on this factor for the Chicano group only. Arithmetic loaded on the Perceptual Organization factor in the White group. In the White sample, Information and Comprehension had significant secondary factor loadings. In the Chicano group, Similarities, Arithmetic, and Comprehension had significant secondary factor loadings.

The second rotated factor, Perceptual Organization, contained Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding for both groups. Digit Span also loaded on this factor for the Anglo group. In the White group, Picture Completion, Block Design, and Object Assembly had significant secondary loadings. Both the Perceptual Organization and Verbal Comprehension factors in this study closely corresponded to those obtained by Reschly (1978) and Kaufman (1975). Coefficients of congruence exceeded .90.

A three-factor solution that was extracted by Gutkin and Reynolds (1980) was not entirely consistent with previous work. The authors cited difficulty interpreting this solution as the second factor split in two. However, this study lent support to the construct validity of Verbal Comprehension and Perceptual Organization factors across Anglo and Chicano ethnic groups.

Dean (1980) compared Anglo (N= 109) and Mexican-American children (N= 123) referred for evaluations in Arizona. Separate principal factor analyses were performed on the 10 standard WISC-R subtests for each group. Varimax rotation followed factor extraction.

Three-factor solutions were obtained for each group. The Verbal Comprehension factor contained Information, Comprehension, Similarities, and Vocabulary for both groups. Picture Arrangement also loaded on this factor for Mexican-American children. Arithmetic had a secondary loading on this factor for both groups. In the Anglo sample, Similarities had a secondary loading on the Perceptual Organization factor. In the Mexican-American sample, Vocabulary and Comprehension had secondary loadings on the Perceptual Organization and Freedom From Distractibility factors, respectively.

The Perceptual Organization factor contained Picture Completion, Picture Arrangement (a secondary loading for the Mexican sample), Block Design, and Object Assembly for both groups. Coding did not load on this factor for either group. In the White group, Block Design had significant loadings on the Verbal Comprehension and Freedom From Distractibility factors as well. In the Mexican sample, Block Design had a secondary loading on Freedom From Distractibility.

The third factor, Freedom From Distractibility, did not resemble the factor obtained in previous work. Arithmetic and Coding had primary loadings on this factor for both groups. The usual subtests loading on this factor are Arithmetic, Digit Span, and Coding. However, Digit Span and Mazes were not included in this analysis.

Dean (1980) indicated coefficients of congruence ranging from .84 to .89 across samples studied. This seems to indicate a fair degree of similarity between factors across groups. Dean (1980) asserted that the Freedom From Distractibility factor is variable when compared with Kaufman (1975).

This study contains notable findings. The first is the dissimilarity of the Freedom From Distractibility factor from previous work. The second is some crossing of verbal subtest loadings on Perceptual Organization factors and performance subtest loadings on Verbal Comprehension factors. It is impossible to determine from this study if these occurrences can only be attributed to the deletion of Digit Span and Mazes. Therefore, this study lends only tentative support for a three- factor solution across Anglo and Mexican-American samples.

Gutkin and Reynolds (1981) performed separate principal factor analyses with varimax rotation for the White (N= 1868) and Black (N= 305) subgroups of the WISC-R standardization sample. Both two- and three-factor solutions were found to fit the data. For the two-factor solution, verbal and performance factors emerged. For both groups, all verbal subtests loaded on the verbal factor and all performance subtests, except Coding, loaded on the performance factor. Coding had a higher loading on the verbal factor for both groups. Several secondary factor loadings were observed for both groups.

The three-factor solution resembled Kaufman's (1975) factors. Verbal Comprehension and Perceptual Organization had their usual loadings. Arithmetic, Digit Span, and Coding comprised the Freedom From Distractibility factor across groups. Again, several secondary loadings were observed.

This study contributes substantially to the factor analytic literature. The samples were large and nationally representative. The only shortcoming is the fact that the WISC-R was not independently standardized within each racial-ethnic group. This point is addressed later.

Part of Sandoval's (1982) study attempted to verify the Gutkin and Reynold's (1980) findings. Principal factor analysis examined intercorrelations of the 12 WISC-R subtests. A Freedom From Distractibility factor emerged for

Anglo (N= 332) children (including Arithmetic's secondary loading) but not for Black (N= 314) or Mexican-American children (N= 307). One may question its meaningfulness as a factor.

Both Verbal Comprehension and Perceptual Organization factors emerged. Information, Similarities, Arithmetic, Vocabulary, and Comprehension loaded on Verbal Comprehension for the Anglo and Black samples. Digit Span replaced Information in the Mexican-American group. Perceptual Organization consisted of Picture Completion, Picture Arrangement, Block Design, and Object Assembly for all groups. In addition, Mazes loaded on this factor for the Black and Mexican-American groups, but not for the Anglo group. Information also loaded on Perceptual Organization for the Mexican group. Again, many secondary factor loadings were observed across groups.

Johnston and Bolen (1984) factor analyzed data generated by referred samples of Black (N= 430) and White (N= 274) children on the 10 mandatory WISC-R subtests. Analysis of two-, three-, and four-factor solutions was accomplished by principal factor analysis. The three-factor solution generated a Verbal Comprehension factor (Information, Similarities, Vocabulary, and Comprehension) and a Perceptual Organization factor (Picture Completion, Picture Arrangement, Block Design, and Object Assembly) for both groups. Arithmetic also loaded on Verbal

Comprehension for the Black group. The Freedom From Distractibility factor contained Arithmetic (secondary loading) and Coding for the Black sample. Arithmetic and Information (secondary loading) comprised this factor for the White sample.

The Verbal Comprehension and Perceptual Organization factors corresponded to Kaufman's (1975) study. However, the Freedom From Distractibility factor exhibited a lower coefficient of congruence. Again, the study supported the verbal-performance dichotomy of the WISC-R. The nature of the Freedom From Distractibility factor remained vague. It is unclear what effect the deletion of Digit Span and Mazes had on the Johnston and Bolen (1984) results.

Juliano, Haddad, & Carroll (1988) performed a principal components analysis with varimax rotation on data generated by White (N= 126) and Black (N= 103) subgroups aged 7-13 years. Eleven subtests, excluding Mazes, were studied. This study assessed stability of the factor structure so this review cites the results of data analysis from the first administration only. Identical three-factor solutions were derived for each group. The verbal factor consisted of Information, Similarities, Vocabulary, and Comprehension. Picture Completion, Picture Arrangement, Block Design, and Object Assembly comprised the performance factor. Freedom From Distractibility was composed of Arithmetic, Digit Span, and Coding. In the Black group, Similarities, Arithmetic, and Coding had secondary loadings on the Freedom From Distractibility, Verbal Comprehension, and Perceptual Organization factors, respectively. In the White sample, Arithmetic, Digit Span, and Block Design had secondary loadings on the Verbal Comprehension factor.

Although Dean (1977) found the WISC-R to be reliable for use with Mexican-American children, Rousey (1990) studied Latin-American (N= 613) and Anglo (N= 668) samples in her attempt to establish factorial validity of the WISC-R Mexicano (WISC-RM). The concern here is not on the Spanish translation and data produced by the Mexican sample. The Latin-American and Anglo WISC-R data is of interest. Intercorrelations of all 12 subtests were analyzed. A combination of principle components and principal factor analysis indicated that a three-factor solution fit WISC-R data for Latino and Anglo samples.

For both groups, Verbal Comprehension consisted of Information, Similarities, Vocabulary, and Comprehension. Arithmetic joined this factor for the Anglo group. The Perceptual Organization factor consisted of Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Mazes for both groups. Freedom From Distractibility consisted of Arithmetic and Digit Span for Latinos and Digit Span and Coding for Anglos. Many subtests had significant secondary factor loadings.

Galkowski, Pietrulewicz, & Scott (1987) appears to be the only study to examine the factor structure of the WISC and WISC-R in a European sample. However, sample size was extremely small (N= 30) and factor analytic results may not be stable. Results indicated the existence of a variation of verbal and performance factors. Freedom From Distractibility did not emerge on the WISC-R. On the WISC, Digit Span and Coding loaded on a third factor. However, Digit Span's highest loading was on the verbal factor.

Summary WISC-R Factor Analytic Research: Ethnic Subgroup Data

Two-factor orthogonal solutions cited above yielded some differences in subtest factor loadings across racial-ethnic subgroups. In addition, in many cases, subtests exhibited significant factor loadings on more than one factor. As mentioned earlier, this summary places a subtest on the factor where it most frequently had its highest factor loading. Tables 6, 7, and 8 are included to present summaries of factor loadings across studies. Note that most studies evidence many instances of secondary subtest factor loadings. This calls into question the usefulness of orthogonal solutions. Table 9 consolidates this information so the reader can more easily see general trends in the factor analyses across studies described above.

As Table 9 shows, the verbal factor for White samples consistently loaded Information, Similarities, Vocabulary, and Comprehension. For Black samples, the verbal factor

Table 6

WISC-R Standardized Subtest Factor Loadings: White Subgroup Data

	1*		2		3	
Subtest	Factor	Loading	Factor	Loading	Factor	Loading
I	VC/PO	.63/.32	VC	.76	VC/PO	.64/.52
S	VC	.59	VC/PO	.68/.39	VC	.83
A	VC/FFD	.43/.45	VC/FFD	.39/.52	VC/PO	.39/.70
V	VC	.74	VC	.76	VC	.79
C	VC	.64	VC	.56	VC/PO	.76/.34
DSp	VC/FFD	.35/.40	--	--	PO	.59
PC	PO	.49	PO	.66	VC/PO	.36/.59
PA	PO	.53	PO	.59	PO	.64
BD	PO	.60	VC/PO/FFD	.31 .73 .47	VC/PO	.38/.51
OA	PO	.59	PO	.69	VC/PO	.35/.59
CD	FFD	.40	PO	.44	PO	.45
MZ	PO	.42	--	--	--	--

Note. *1=Reschly (1978) Three-factor solution; 2=Dean (1980) Three-factor solution; 3=Gutkin & Reynolds (1980) Two-factor solution. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; VC=Verbal Comprehension; PO=Perceptual Organization; FFD=Freedom From Distractibility; --=subtest not examined.

Table 6 (continued)

WISC-R Standardized Subtest Factor Loadings: White Subgroup Data

	4a	b*	a	b	5		6	
Subtest	Factor		Loading		Factor	Loading	Factor	Loading
I	V	V/F	.72	.63/.35	V/F	.67/.36	V/F	.54/.33
S	V/PO	V/PO	.67/.33	.63/.32	V	.61	V	.71
A	V	V/F	.58	.37/.55	V/F	.53/.37	F	.98
V	V	V/F	.81	.77/.31	V	.80	V	.65
C	V	V	.62	.64	V	.62	V	.64
DSp	V	F	.43	.60	F	.73	--	--
PC	PO	V/PO	.53	.31/.53	PO	.48	PO	.45
PA	V/PO	V/PO	.33/.45	.32/.44	V/PO	.38/.42	PO	.42
BD	V/PO	PO/F	.33/.73	.72/.30	PO	.64	PO	.54
OA	PO	PO	.67	.66	PO	.77	PO	.70
CD	V	F	.31	.36	F	.30	NS	NS
MZ	PO	PO	.45	.44	PO/F	.32/.31	--	--

Note. *4=Gutkin & Reynolds (1981), a=Two-factor solution, b=Three-factor solution; 5=Sandoval (1982) Three-factor solution; 6=Johnston & Bolen (1984) Three-factor solution. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; VC=Verbal Comprehension; PO=Perceptual Organization; F=Freedom From Distractibility; --=subtest not examined; NS=factor loading < .30.

Table 6 (continued)

WISC-R Standardized Subtest Factor Loadings: White Subgroup Data

	6[*]		7	
Subtest	Factor	Loading	Factor	Loading
I	VC	.74	VC/FFD	.59/.36
S	VC	.81	VC	.64
A	VC/FFD	.47/.56	VC/FFD	.48/.39
V	VC	.84	VC	.72
C	VC	.71	VC	.61
DSp	VC/FFD	.33/.66	VC/FFD	.31/.45
PC	PO	.80	VC/PO	.45/.47
PA	PO	.64	VC/PO	.34/.40
BD	VC/PO	.30/.70	PO	.60
OA	PO	.76	PO	.65
CD	FFD	.80	FFD	.32
MZ	--	--	PO/FFD	.38/.32

Note. [*]6=Juliano et al. (1988) Three-factor solution; 7=Rousey (1990) Three-factor solution. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; VC=Verbal Comprehension; PO=Perceptual Organization; FFD=Freedom From Distractibility; -- =subtest not examined.

Table 7

WISC-R Standardized Subtest Factor Loadings: Black Subgroup Data

	1*		2		3a	3b	3a	3b
Subtest	Factor	Loading	Factor	Loading	Factor		Loading	
I	V	.55	V/PO	.66/.44	V	V/F	.72	.58/.42
S	V	.36	V/PO	.59/.41	V/PO	V	.66/.30	.62
A	V	.37	V/PO	.61/.34	V	V/F	.63	.34/.68
V	V	.67	V	.75	V	V/F	.81	.80/.33
C	V	.40	V	.71	V/PO	V/PO	.59/.34	.58/.33
DSp	--	--	V/F	.49/.36	V	F	.49	.57
PC	PO	.33	PO	.52	V/PO	V/PO	.36/.56	.38/.55
PA	PO	.56	PO	.53	V/PO	PO	.30/.47	.47
BD	PO	.47	PO/F	.33/.58	V/PO	PO/F	.32/.68	.67/.32
OA	PO	.51	F	.58	PO	PO	.73	.72
CD	None	NS	V	.33	V	None	.30	NS
MZ	--	--	PO/F	.34/.30	PO	PO	.54	.53

Note. *1=Vance & Wallbrown (1978) Two-factor solution; 2=Reschly (1978) Three-factor solution; 3=Gutkin & Reynolds (1981), a=Two-factor solution, b=Three-factor solution. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; VC=Verbal Comprehension; PO=Perceptual Organization; F=Freedom From Distractibility; --=subtest not examined; NS=factor loading < .30.

Table 7 (continued)

WISC-R Standardized Subtest Factor Loadings: Black Subgroup Data

	4*		5		6	
Subtest	Factor	Loading	Factor	Loading	Factor	Loading
I	V	.73	V	.48	V	.48
S	V	.61	V	.50	V/PO	.67/.33
A	V/PO	.53/.37	V/F	.44/.41	V/F	.45/.52
V	V	.71	V	.77	V	.86
C	V	.71	V	.65	V	.75
DSp	V	.30	--	--	F	.71
PC	V/PO	.39/.41	PO	.42	PO	.63
PA	V/PO	.34/.53	PO	.37	PO	.70
BD	PO	.65	PO	.54	PO	.77
OA	PO	.65	PO	.66	PO	.64
CD	PO	.32	F	.39	PO/F	.31/.65
MZ	PO	.54	--	--	--	--

Note. *4=Sandoval (1982) Two-factor solution; 5=Johnston & Bolen (1984) Three-factor solution; 6=Juliano et al. (1988) Three-factor solution. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; VC=Verbal Comprehension; PO=Perceptual Organization; F=Freedom From Distractibility; -- =subtest not examined.

Table 8

WISC-R Standardized Subtest Factor Loadings: Hispanic Subgroup Data

	1*		2		3	
Subtest	Factor	Loading	Factor	Loading	Factor	Loading
I	V/F	.66/.33	V	.80	V	.71
S	V	.67	V	.76	V/PO	.69/.33
A	V/F	.40/.45	V/F	.58/.63	V/PO	.48/.31
V	V/F	.67/.30	V/PO	.85/.31	V	.70
C	V	.61	V/F	.46/.30	V/PO	.55/.33
DSp	V/F	.33/.31	--	--	None	NS
PC	V/PO	.32/.52	PO	.54	PO	.56
PA	PO/F	.38/.39	V/PO	.48/.44	PO	.72
BD	PO	.59	PO/F	.62/.39	PO	.64
OA	PO	.58	PO	.79	PO	.51
CD	F	.37	F	.38	PO	.37
MZ	PO	.47	--	--	--	--

Note. *1=Reschly (1978) Three-factor solution; 2=Dean (1980) Three-factor solution; 3=Gutkin & Reynolds (1980) Two-factor solution. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; VC=Verbal Comprehension; PO=Perceptual Organization; F=Freedom From Distractibility; --=subtest not examined; NS=factor loading $\leq .30$.

Table 8 (continued)

WISC-R Standardized Subtest Factor Loadings: Hispanic Subgroup Data

	4[a]		5		6	
Subtest	Factor	Loading	Factor	Loading	Factor	Loading
I	PO	.38	V	.84	V/PO/F	.59/.31/.40
S	V	.72	V	.77	V	.64
A	V/PO	.54/.39	V/F	.66/.40	V/PO/F	.34/.30/.55
V	V	.81	V/PO	.82/.30	V	.77
C	V	.62	V	.82	V	.66
DSp	V/PO	.48/.34	V/F	.38/.42	V/PO/F	.34/.32/.42
PC	PO	.58	PO	.83	PO	.45
PA	V/PO	.38/.41	V/PO	.48/.31	PO/F	.36/.32
BD	PO	.60	PO	.75	PO	.59
OA	PO	.54	PO/F	.68/.36	PO	.54
CD	PO	.32	F	.94	PO	.31
MZ	PO	.48	--	--	PO	.50

Note. [a]4=Sandoval (1982) Two-factor solution; 5=Stedman et al. (1978) Three-factor solution. 6=Rousey (1990) Three-factor solution. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; V=Verbal Comprehension; PO=Perceptual Organization; F=Freedom From Distractibility; --=subtest not examined.

Table 9

WISC-R Factor Structure Based on Literature Review of White, Black, and Hispanic Samples: Two-Factor Solution

Verbal Factor		
White	Black	Hispanic
Information	Information	Similarities
Similarities	Similarities	Arithmetic
Vocabulary	Arithmetic	Vocabulary
Comprehension	Vocabulary	Comprehension
	Comprehension	
	Digit Span	

Performance Factor		
White	Black	Hispanic
Picture Completion	Picture Completion	Picture Completion
Picture Arrangement	Block Design	Picture Arrangement
Block Design	Object Assembly	Block Design
Object Assembly	Mazes	Object Assembly
Mazes		Coding
		Mazes

consisted of Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span. For Hispanic samples, the verbal factor consisted of Similarities, Arithmetic, Vocabulary, and Comprehension. Note the differences across racial-ethnic groups on this factor.

Table 9 displays the performance factor for each group as well. For White samples, the performance factor consisted of Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Mazes. For Black samples, Picture Completion, Block Design, Object Assembly, and Mazes comprised the performance factor. The Hispanic performance factor consisted of Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, and Mazes.

Each group contained one or more subtests which did not load consistently on verbal or performance factors. Coding did not load consistently on any one factor across studies of Black samples (Vance & Wallbrown, 1978; Gutkin & Reynolds, 1981; Sandoval, 1982). Across studies of Hispanic samples (Gutkin & Reynolds, 1980; Sandoval, 1982) Information and Digit Span did not load consistently on any particular factor. Finally, across studies of White samples (Reschly, 1978; Gutkin & Reynolds, 1980; Gutkin & Reynolds, 1981) Arithmetic, Digit Span, and Coding did not load consistently on any particular factor. These subtests comprise the Freedom From Distractibility factor in three-factor solutions.

Orthogonal three-factor solutions were more consistent across racial-ethnic subgroups as Table 10 shows. For all subgroups, the Verbal Comprehension factor included Information, Comprehension, Similarities, and Vocabulary. Picture Arrangement also loaded consistently on this factor for the Hispanic subgroup only. The Perceptual Organization factor consisted of Picture Completion, Picture Arrangement (except for the Hispanic group), Block Design, Object Assembly, and Mazes. Finally, the Freedom From Distractibility factor contained Arithmetic, Digit Span, and Coding for White and Hispanic groups. Only Digit Span loaded consistently on Freedom From Distractibility for Black samples.

It appears that the orthogonal three-factor solution for the <u>WISC-R</u> represents Wechsler's conceptualization of intelligence fairly well across racial-ethnic subgroups. With the exception of Picture Arrangement's verbal loading for the Hispanic subgroup, factor content of Verbal Comprehension and Perceptual Organization was equivalent across racial-ethnic subgroups. Support for Freedom From Distractibility was evidenced for the White and Hispanic groups.

Caution must be exercised when interpreting these general trends. First, within individual studies, subtests loading .30 or better on more than one factor was more of the rule than the exception. Second, analysis of general

Table 10

WISC-R Factor Structure Based on Literature Review of White, Black, and Hispanic Samples: Three-Factor Solution

Verbal Factor		
White	Black	Hispanic
Information	Information	Information
Similarities	Similarities	Similarities
Vocabulary	Vocabulary	Vocabulary
Comprehension	Comprehension	Comprehension
		Picture Arrangement
Perceptual Organization Factor		
White	Black	Hispanic
Picture Completion	Picture Completion	Picture Completion
Picture Arrangement	Picture Arrangement	Block Design
Block Design	Block Design	Object Assembly
Object Assembly	Object Assembly	Mazes
Mazes	Mazes	
Freedom From Distractibility Factor		
White	Black	Hispanic
Arithmetic	Digit Span	Arithmetic
Digit Span		Digit Span
Coding		Coding

trends across studies may obscure departures from expected factor loadings within individual studies. Third, despite the promise of consistent factor content across racial-ethnic subgroups, all analyses were done on scaled scores. Scaled scores represent an overwhelming contribution of data generated by White children which may mask unique response patterns of minority subgroups.

Factor Analyses of the WISC-III: Standardization Data

The standardization sample for the WISC-III included 100 boys and 100 girls at each age level from 6-16 years inclusive. The sample was stratified across age, gender, race-ethnicity, geographic region, and parent education. The sample approximated the 1988 United States' census data.

Initial studies of the WISC-III factor structure appear in the WISC-III manual (Wechsler, 1991). These initial studies were important since a new subtest, Symbol Search, was added to the test battery. The manual reports results of both exploratory and confirmatory factoring techniques. Both sets of data analyses were conducted on the well-stratified standardization sample. Factor analyses were performed for four age ranges: 6-7, 8-10, 11-13, and 14-16. Results consistently suggest the presence of four factors (Wechsler, 1991).

The manual reports studies employing maximum likelihood exploratory factoring techniques with varimax rotation. Specification of a two-factor solution results in a bifurcation between verbal and performance subtests across the age range. Specification of a four-factor solution generally supported: Verbal Comprehension, Perceptual Organization, Freedom From Distractibility, and Processing Speed factors.

Inconsistency in factor composition is evident. The Verbal Comprehension factor consistently contained Information, Similarities, Vocabulary, and Comprehension. The Perceptual Organization factor contained Picture Completion (except at ages 6 and 7), Picture Arrangement, Block Design, Object Assembly, and Mazes (except ages 6 and 7). Freedom From Distractibility contained Arithmetic and Digit Span across the age range. However, Mazes, Picture Arrangement, and Symbol Search loaded significantly at ages 6 and 7. Processing Speed contained Coding and Symbol Search (except ages 6 and 7).

Sattler (1992) argued that evidence from his own work does not support the existence of Freedom From Distractibility. Eigenvalues of less than 1.0 emerged for the factor. Sattler's orthogonal three-factor solution contained: Verbal Comprehension (Information, Similarities, Vocabulary, and Comprehension); Perceptual Organization (Block Design, Object Assembly, and Picture Completion); and Processing Speed (Coding and Symbol Search). In Sattler's analysis, many subtests are left out of his three-factor solution or had significant loadings on more than one factor. Similarities and Arithmetic had loadings exceeding .30 on Verbal Comprehension and Perceptual Organization. Digit Span loaded significantly on Verbal Comprehension, but Sattler did not include the subtest on this factor. Picture Completion, Picture Arrangement, and Block Design all had significant loadings on Verbal Comprehension. Symbol Search and Mazes both had significant loadings on Perceptual Organization.

Roid, Prifitera, and Weiss (1993) deemed the eigenvalue an arbitrary guideline. They add that Sattler's investigation of factor structure at each age of the standardization sample resulted in data analyses of smaller sample sizes. Hence, his results may be affected by sampling fluctuation (Roid et al., 1993).

Confirmatory factor analysis tested several models containing one to five factors. Wechsler (1991) reported the use of several indices of model fit which included chi-square/df, the Tucker-Lewis Index, Adjusted Goodness of Fit Index, and Root Mean Square Residual. Chi-square difference analyzed successive improvement in model fit. Results again suggested a four-factor solution.

Roid et al. (1993) re-examined WISC-III factor structure in an independent nationally representative sample. Several exploratory factor analyses examined subtest intercorrelations. Maximum-likelihood chi-square statistics indicated the presence of five factors. The fifth was dubbed a "singleton" factor which was the Comprehension subtest. Roid et al. (1993) argued cogently that this splitting from the Verbal Comprehension factor was not sufficient evidence to consider a five-factor solution.

Maximum likelihood confirmatory factor analysis tested a variety of models, including one with the singleton fifth factor. Like Wechsler (1991), Roid et al. (1993) used a number of model-fit indices and again the four-factor solution was considered most plausible. Verbal Comprehension consisted of Vocabulary, Similarities, Information, and Comprehension. Perceptual Organization consisted of Object Assembly, Block Design, Picture Completion, and Picture Arrangement. Freedom From Distractibility consisted of Arithmetic and Digit Span. Processing Speed loaded Symbol Search and Coding.

Kamphaus, Hutchinson, and Platt (1994) investigated the Wechsler (1991) four-factor model as well as two other competing models: Wechsler's original two-factor model (excluding Symbol Search) and Kaufman's three-factor model (also excluding Symbol Search). Subtest intercorrelations at each age level of the standardization sample were analyzed. Factor correlations were estimated for each model.

Results suggested little difference between the three- and four-factor solutions at ages 6 and 9. For all other age levels, none of the models fit very well. However, Kamphaus et al. (1994) added that the four-factor model consistently evidenced the best improvement in fit over the one-factor baseline model as indicated by the Tucker-Lewis Index. The authors noted significant factor intercorrelations and questioned the use of orthogonal rotations in WISC-III research.

Reynolds and Ford (1994) replicated Kaufman's (1975) study. Data from each age level of the standardization sample were analyzed. Only the 12 subtests common to the WISC-R and WISC-III were included (Symbol Search was not included). Principal components and principal factor analysis generated three interpretable factors: Verbal Comprehension, Perceptual Organization, and Freedom From Distractibility.

Allen & Thorndike (1995) combined exploratory and confirmatory procedures in their study of WISC-III and the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R; Wechsler, 1989) factor structure. Two factors were specified for extraction (verbal and performance) since these factors are generally supported most often in the literature (Allen & Thorndike, 1995). The study confirmed these factors across tests.

WISC-III Factor Analytic Research: Racial-Ethnic Subgroups

The only study reviewed assessing WISC-III factor structure within racial-ethnic subgroups was by Slate & Jones (1995). The investigated WISC-III factor structure in

58 African-American boys aged 6-16 referred for psychological evaluations. Only 11 of 13 subtests were analyzed (Mazes and Symbol Search deleted). Unrotated principal components procedures were performed three times: (a) once on the 10 core subtests plus Digit Span; (b) once on the Verbal Scale subtests plus Digit Span only; and (c) once on the Performance Scale subtests only. Results yielded support for general, verbal, and performance factors. However, when factors were subjected to varimax rotation, the factors did not resemble those in previous work. These rotated factors demonstrated an intermixing of verbal and performance subtests.

Summary of WISC-III Factor Analytic Research

Wechsler's (1991) exploratory analysis of an orthogonal two-factor solution to WISC-III standardization data seems to support the Verbal and Performance Scale organization of the test. It must be noted that Information, Arithmetic, and Digit Span had factor loadings exceeding .30 on the performance factor. Picture Completion, Picture Arrangement, and Block Design had significant loadings on the verbal factor. Again, this calls the use of orthogonal solutions into question. As noted above, Allen & Thorndike (1995) found support for a two-factor solution.

Table 11 shows results of four-factor solutions by Wechsler (1991) and Roid et al. (1993). Results generally support the four-factor solution cited in Wechsler (1991).

Table 11

WISC-III Standardized Factor Loadings: Standardization Sample Data

	1[1]		2[2]	
Subtest	Factor	Loading	Factor	Loading
I	VC	.72	VC	.72
S	VC	.72	VC	.75
A	VC/FFD	.41/.73	VC/FFD	.49/.53
V	VC	.79	VC	.81
C	VC	.65	VC	.68
DSp	FFD	.34	FFD/PS	.35/.33
PC	VC/PO	.38/.53	VC/PO	.40/.50
PA	VC/PO	.33/.37	VC/PO	.33/.43
BD	PO	.70	PO	.67
OA	PO	.69	PO	.72
CD	PS	.79	PS	.56
MZ	PO	.36	PO	.41
SS	PO/PS	.35/.56	PS	.71

Note. 1=Wechsler (1991) Four-factor solution; 2=Roid et al. (1993) Four-factor solution. I=Information; C=Comprehension; A=Arithmetic; S=Similarities; V=Vocabulary; DSp=Digit Span; PC=Picture Completion; PA=Picture Arrangement; BD=Block Design; OA= Object Assembly; CD=Coding; MZ=Mazes; SS=Symbol Search; VC=Verbal Comprehension; PO=Perceptual Organization; FFD=Freedom From Distractibility; PS=Processing Speed.

The verbal factor consistently contained Information, Comprehension, Similarities, and Vocabulary. The performance factor contained Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Mazes. Freedom From Distractibility generally loaded Digit Span and Arithmetic. Processing Speed contained Coding and Symbol Search. As Table 11 indicates, many secondary loadings were obtained. This is consistent with previous orthogonal exploratory analyses of the WISC-R.

Confirmatory factor analyses generally support a four-factor solution to the WISC-III. Wechsler (1991) and Roid et al. (1993) provide evidence for the four factors. Kamphaus et al. (1994) suggested a four-factor solution was probably the best fit for models assessed but the fit was not necessarily good. Allen & Thorndike (1995) preferred a verbal-performance two-factor solution.

A four-factor solution typically emerges when analyzing covariance matrices of sample sizes exceeding 200. Analysis of each separate age level of the standardization sample results in poorer fit of the four-factor model.

At this time, no factor analytic study has examined WISC-III factor structure across several racial-ethnic minority groups. This study examines structural and measurement invariance of the WISC-III across White, Black, and Hispanic subgroups of the standardization sample. Each subgroup is nationally representative based on 1988 United

States' census data.

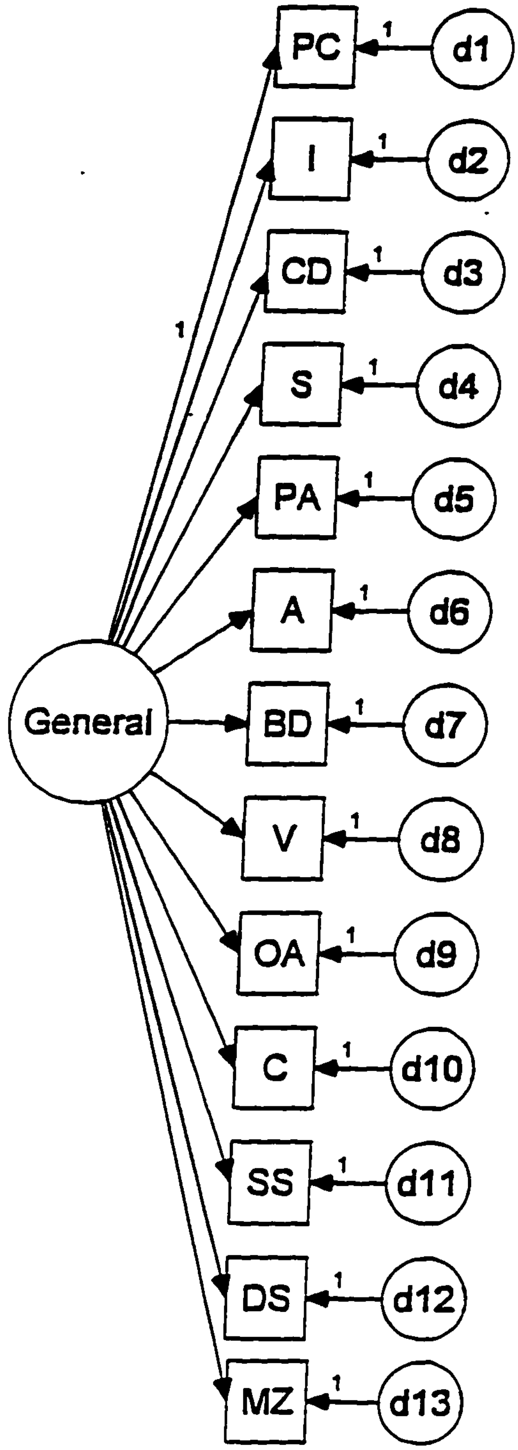
Hypotheses

Analyses of WISC-R and WISC-III factor solutions above helped shape the present hypotheses. Most empirical work on Wechsler's tests suggest multifactor models. Three-factor solutions of the WISC-R (Verbal Comprehension, Perceptual Organization, and Freedom From Distractibility) appear to be more slightly more consistent across racial-ethnic subgroups, with respect to subtest factor loadings, than two-factor solutions. However, Freedom From Distractibility was a source of controversy with the WISC-R and remains so with the WISC-III (Sattler, 1992). At this time, it is unclear how the new Processing Speed factor will fare in studies of racial-ethnic subgroups.

In this study, it was hypothesized that the WISC-III would exhibit partial factorial invariance through analyses of subtest raw and standard score data across the White, Black, and Hispanic subgroups of the standardization sample. This means that factor models across racial-ethnic subgroups are not exactly alike. Factor models assessed were: (a) a one-factor model (Figure 1); (b) a correlated two-factor model (Figure 2); (c) a three-factor model which, in this study, allowed for correlated factors (Figure 3); and (d) a four-factor model which, in this study, also allowed for correlated factors (Figure 4).

Three hypotheses were tested in this study. Hypothesis

Figure 1 One-Factor Model*



***Note.** PC = Picture Completion; I = Information; CD = Coding; S = Similarities; PA = Picture Arrangement; A = Arithmetic; BD = Block Design; V = Vocabulary; OA = Object Assembly; C = Comprehension; SS = Symbol Search; DS = Digit Span; MZ = Mazes; d = Disturbance

Figure 2 Two-Factor Model

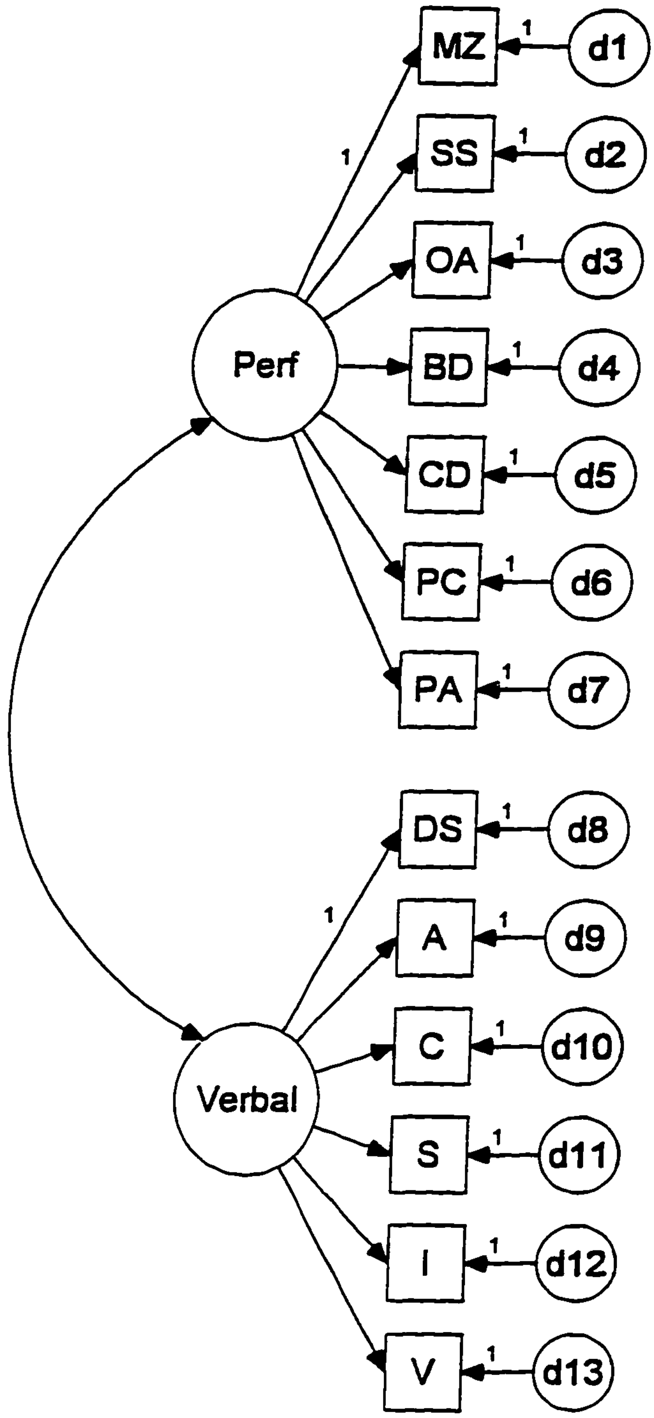


Figure 3 **Three-Factor Model**

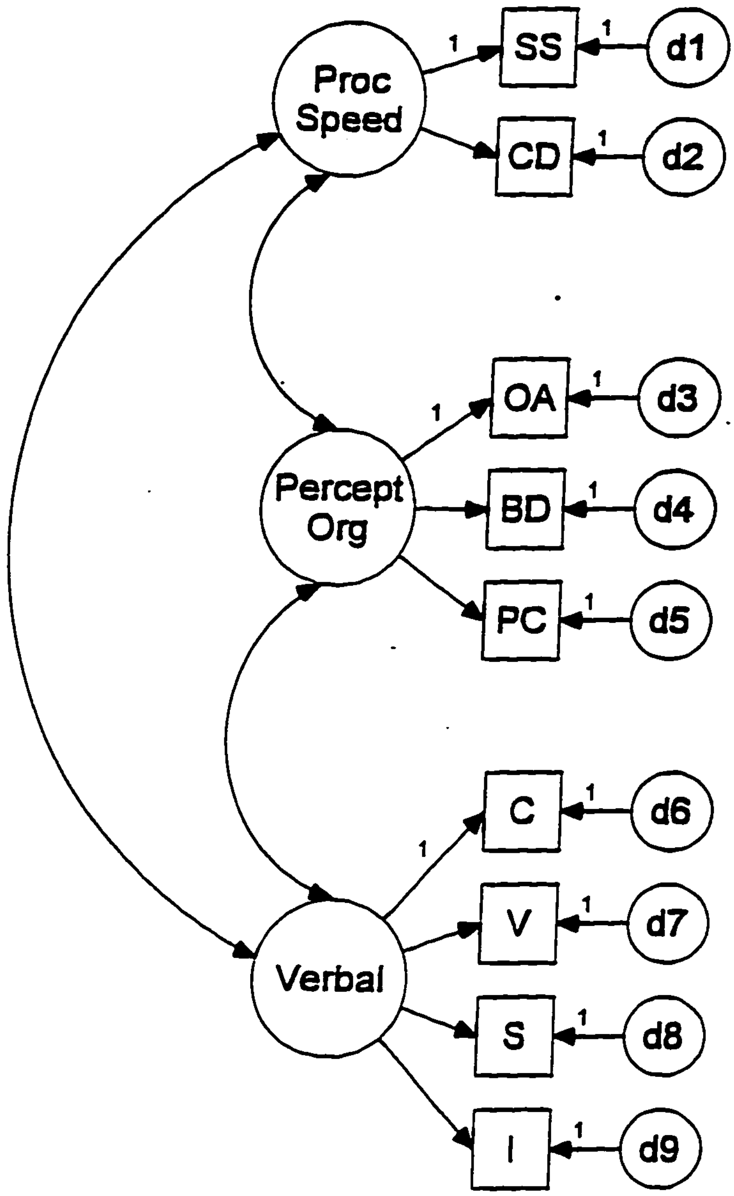
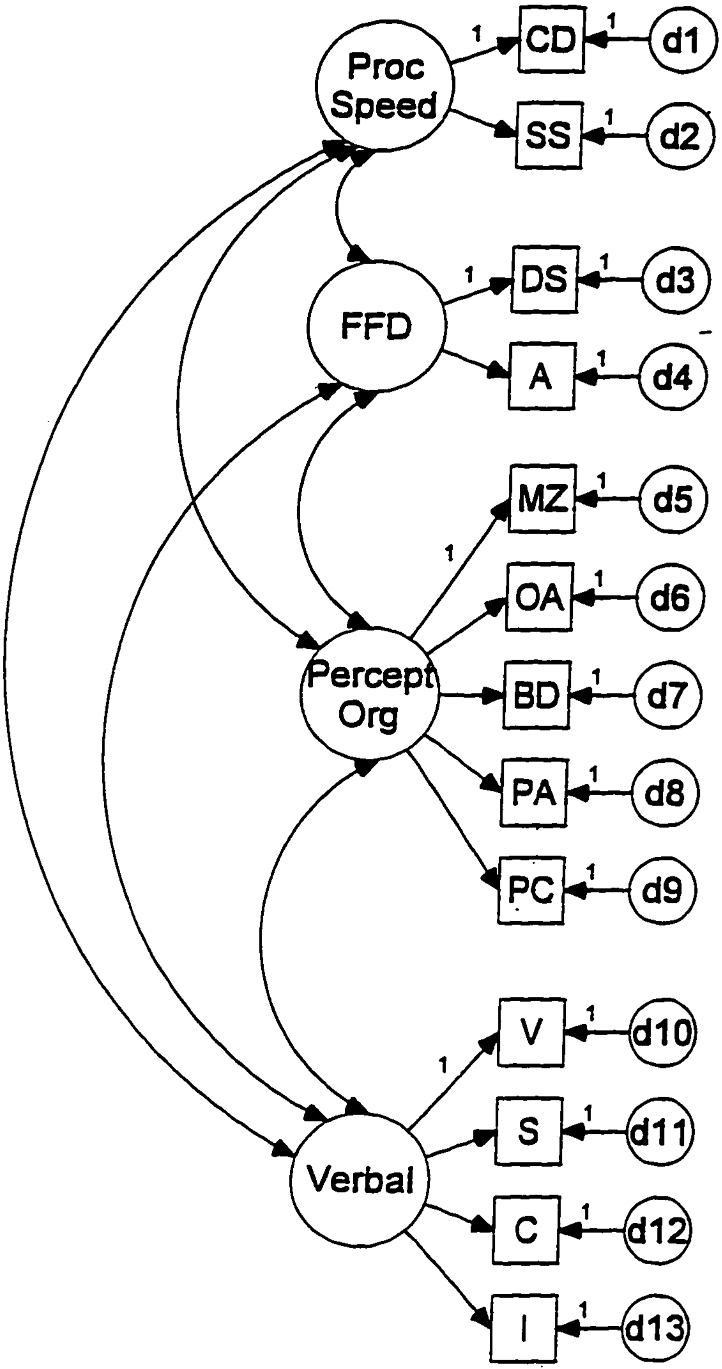


Figure 4 Four-Factor Model



1 stated that the unconstrained correlated two-factor model (verbal and nonverbal factors) best fit the WISC-III subtest variance-covariance matrices for all groups. This is tantamount to saying that two factors best explain the variance-covariance matrices for each racial-ethnic subgroup. One-, three-, and four-factor solutions would not fit the data as well. Hypothesis 2 stated that the pattern of factor loadings (i.e., content of the two factors) would differ across groups. It was expected that these two factors would retain most of their respective verbal and nonverbal properties within each racial-ethnic group. Different combinations of subtests would comprise these factors across groups. For those subtests that consistently loaded on the same factor across groups, their relative rank orders would differ. Figure 2 illustrates the two-factor model which defines the organization of the WISC-III. However, Figure 2 was expected to be modified based on the present results. Such a modification will more accurately reflect group-specific models. Hypotheses 1 and 2 were expected to hold for analyses of both subtest scaled score and raw score variance-covariance matrices.

Hypothesis 3 stated that analysis of raw score data within each racial-ethnic subgroup would reveal idiosyncratic response patterns. Results from raw score analyses would show greater discrepancies between groups than analyses of scaled score data. These discrepancies would appear in both factor loading patterns and relative rank ordering of factor loadings.

Chapter II

METHOD

Data

The data analyzed in this study were subtest variance-covariance matrices were prepared by The Psychological Corporation. Six WISC-III subtest variance-covariance matrices were generated for analysis. One scaled score matrix and one raw score matrix was created for each racial-ethnic subgroup. All 13 subtests were included for the present analyses.

Separate matrices reflected the response patterns made by the White (N= 1543), Black (N= 338), and Hispanic (N= 242) subgroups of the WISC-III standardization sample. Children aged 6-16 years were categorized in the White or Black subgroups based on their parents' race-ethnic group membership (Wechsler, 1991). In addition, Wechsler (1991) indicated that children were assigned to the Hispanic group if parents indicated that a child was of Hispanic origin. This assignment was made regardless of how the parents reported their own racial-ethnic identity. The entire standardization sample was stratified along the variables of age, gender, race-ethnicity, geographic region, and parent education.

Wechsler (1991) illustrates sample representation from each category and demonstrates a close match with the 1988 United States' census. Analyses of these White, Black, and

Hispanic racial-ethnic subgroups allowed for more valid generalizations of the present results their respective populations. Previous studies of the WISC, WISC-R, and WISC-III did not analyze data from nationally representative racial-ethnic subgroups.

Procedure

Variance-covariance matrices of the 13 subtest scaled scores and raw scores were input into Amos 3.61 (Arbuckle, 1997). Within groups analyses and tests for factorial invariance were applied to both scaled score and raw score data. The procedure was performed for the four models described above: (a) the one-factor model (Figure 1); (b) the correlated two-factor model (Figure 2); (c) the correlated three-factor model (Figure 3); and (d) the correlated four-factor model (Figure 4).

Several within groups confirmatory factor analyses were performed for each racial-ethnic subgroup across the four models and datasets. These analyses helped assess Hypothesis 3 described above. Within groups analyses also determined if each model fit the data well within each group before proceeding to the simultaneous analysis. Further, the within group analyses allowed for: (a) comparisons between competing models; and (b) comparisons of results between datasets for each group separately. Various fit indices were used to evaluate each model: an overall chi-square statistic ($\underline{X}^2$), chi-square/df ($\underline{X}^2$/df), the Tucker-

Lewis Index (TLI; Tucker & Lewis, 1973), and the Adjusted Goodness of Fit Index (AGFI).

The $\underline{X}^2$ test assesses whether residual differences between the original variance-covariance matrix and the matrix reproduced by the model are significant. A significant $\underline{X}^2$ indicates that the model inadequately reproduced or "fit" the original matrix. Alternative models for the same data which estimate more parameters generate lower $\underline{X}^2$ values. The seemingly better fit of these models may be statistical artifact. The $\underline{X}^2$/df ratio incorporates a penalty function for using more parameters in a model (more parameters means less df). Ratios ranging anywhere from 2 to 5 may indicate acceptable fit (Bollen, 1989).

The TLI indicates the percentage of covariance explained by the model in question. Values of less than .90 indicate that the model can be improved upon (Bentler & Bonett, 1980). The AGFI is another index that incorporates a penalty function for additional parameters and indicates the relative amount of variances and covariances explained by the model. Values .90 or greater indicate acceptable fit.

Marsh, Balla, & McDonald (1988) indicated that each fit index was influenced by sample size. Both $\underline{X}^2$ and $\underline{X}^2$/df varied with sample size for false models. They noted that the Goodness of Fit Index (GFI) was best for stand alone indices and that the AGFI often overcorrected for sample size. In this study, the AGFI may represent a conservative test of factor models. The TLI was the best index when considering incremental fit (a target model over a baseline model). Given these findings, it was imperative to use several commonly used fit indices to generate a pattern of evidence leading to the acceptance or rejection of a particular model. Exclusive reliance on one index may have been misleading.

In this study, when $\underline{X}^2$ values conflicted with TLI, AGFI, or even $\underline{X}^2$/df, more confidence was placed in the latter three indices when determining the fate of a model. The TLI and AGFI indices provide more specific quantitative information in terms of percentages of how well a model fits the data. More confidence was placed in models that had the most number of favorable fit indices.

The procedure used to test invariance was one advocated by Bollen (1989) and discussed by Byrne, Shavelson, & Muthen (1989). For each factor model, a set of increasingly restrictive hypotheses were tested. The hypotheses tested were: (a) the number of factors of the proposed model provided an equally good fit to the data across groups; (b) the factor loading pattern was the same across groups; (c) subtests were measured with the same reliability across groups (invariance of error); (d) factor variances were equivalent across groups; and (e) factor covariances were equivalent across groups.

To illustrate, the one-factor model was tested simultaneously across all groups without specifying that any part of the model be equal across groups. This is tantamount to saying that the one-factor model was left unconstrained across groups. Next, the factor loadings were forced to be equal across groups (i.e., factor loadings were constrained). As the test for invariance proceeded, more constraints were added: error variances were forced to be equal across groups and finally the factor variances. Assessment of factor covariances was not relevant to the one-factor model.

As in the within groups analyses $\underline{X}^2$, $\underline{X}^2$/df, TLI, and AGFI were used to evaluate the number of factors hypothesis. Tests for invariance of factor loadings, error variances, factor variances and covariances included additional fit indices: chi-square change ($\Delta\underline{X}^2$), and chi-square change/df change ($\Delta\underline{X}^2/\Delta$df). These indices evaluated whether a more constrained model provided for a better fit than a less constrained model. A significant $\Delta\underline{X}^2$ indicated a poorer fit of the more constrained model relative to the less constrained model. $\Delta\underline{X}^2/\Delta$df indicated acceptable fit of the more constrained model if the value was less than 5. Since both change indices are affected by sample size, all indices mentioned were considered together to evaluate models.

The final part of the study evaluated relative rank orders of factor loadings across groups for similar factors. This procedure was done for both scaled and raw score datasets to determine if common factors had similar qualitative emphasis across groups. Results from each group were compared to one another and were also compared to results from the simultaneous analysis. This is important since results from simultaneous analyses are influenced by characteristics of the study's largest subgroup.

Chapter III

RESULTS

Within Groups Analyses: General Findings

Within groups analyses of scaled score and raw score data suggested that the four-factor model was the best fitting and most generalizable model across the racial-ethnic subgroups. Support for the three-factor model was found for the Black and Hispanic groups. The three-factor model did not fit the White group's data as well.

It is difficult to compare results from the three- and four-factor models since different numbers of subtests are used in their composition. In both models all subtests evidenced significant factor loadings across groups. However, analysis of the White group scaled and raw score data did not provide as convincing support for the three-factor model as it did for the four-factor model. Most notably, $\underline{X}^2$/df indices were not acceptable for either dataset. Since TLI and AGFI were acceptable across datasets one could not conclude that the three-factor model represents a bad fit for White data. However, it is not as good as the fit of the four-factor model which is discussed below. One would expect a better fit for the three-factor model given that it uses 9 observed variables compared to the four-factor model's 13. Therefore, the three-factor model appears less generalizable than the four-factor model. Results regarding the three-factor model are presented in

Appendices A through H. The rest of this section focuses on the one-, two-, and four-factor models since they are each comprised of 13 subtests and direct comparisons are possible.

Scaled Score Data

The within groups analysis determined which models, if any, fit the data for each group separately. It also allowed for factor loading comparisons to the simultaneous analyses which are presented later. Table 12 presents results of within groups confirmatory factor analyses performed on scaled score data. All factor models were tested for fit in each of the racial-ethnic groups separately. Inspection of fit indices suggest that, for models using all 13 WISC-III subtests, the four-factor model fit the data best for each racial-ethnic subgroup. Results did not support Hypothesis 1 that a two-factor solution would best fit the data for each group.

Except for the X^2 test, indices of fit were favorable for the four-factor model within each ethnic group. Each X^2 test was statistically significant across groups. This suggests that the model does not fit the data well. However, the X^2 statistic is sensitive to sample size. Note that the value of X^2 varies with sample size in this study. For Whites, the largest sample (N=1543), the obtained X^2 was 233.55. For Hispanics, the smallest sample (N=242), the obtained X^2 was 84.03. This pattern also held for values of

Table 12

Within Groups Confirmatory Factor Analyses of Scaled Score Data

One-Factor Model	Chi-Square	df	p	Chi-Square/df	TLI	AGFI
White	1311.76	65	<.001	20.19	.79	.81
Black	308.84	65	<.001	4.75	.82	.80
Hispanic	141.58	65	<.001	2.18	.90	.87
Two-Factor Model	**Chi-Square**	**df**	**p**	**Chi-Square/df**	**TLI**	**AGFI**
White	738.51	64	<.001	11.54	.88	.90
Black	184.87	64	<.001	2.89	.91	.88
Hispanic	99.59	64	<.001	1.56	.95	.91
Four-Factor Model	**Chi-Square**	**df**	**p**	**Chi-Square/df**	**TLI**	**AGFI**
White	233.55	59	<.001	3.96	.97	.96
Black	105.19	59	<.001	1.78	.96	.93
Hispanic	84.03	59	<.001	1.42	.97	.92

$\underline{X}^2$/df which is also sensitive to sample size.

Inspection of the other fit indices consistently indicated adequate model fit for the four-factor model. For each group, the $\underline{X}^2$/df index was less than 5.0. TLI and AGFI, values exceeded .90 for each group which indicated good model fit. The TLI was .97 for White and Hispanic groups and .96 for the Black group. This means that the four-factor model accounts for 97 percent of covariances in the White and Hispanic groups and 96 percent of covariances in the Black group. The AGFI indicated that the relative amount of variances and covariances explained by the four-factor model was adequate. The model accounted for 96 percent of White group (AGFI=.96), 93 percent of Black group (AGFI=.93), and 92 percent of Hispanic group (AGFI=.92) variances and covariances.

For the one- and two-factor models, one or more fit indices indicated inadequate fit for scaled score data matrices. The one-factor model fared poorly for the White and Black group on all fit indices except for $\underline{X}^2$/df for the Black group. The one-factor model did somewhat better in the Hispanic sample. However, the AGFI was below the .90 criterion and the TLI was borderline at .90.

Table 12 demonstrates better fit for the two-factor model as compared to the one-factor model for Black and Hispanic groups. However, the AGFI for the Black group failed to meet the .90 criterion. For the Hispanic sample, each fit index indicated adequate fit. All fit indices for the Hispanic four-factor model were superior to the two-factor model.

Raw Data

The pattern of results seen in the analysis of raw data is similar to the one seen in the scaled score data analysis. Table 13 presents fit indices for each model based on raw data across each racial-ethnic group. Once again, results indicated strong support for four factors. Like the within groups analysis of scaled score data, within groups analysis of raw data did not support the hypothesis that a two-factor model would explain the data best for each group.

With the exception of $\underline{X}^2$ tests and $\underline{X}^2$/df for the White group ($\underline{X}^2$/df=6.69), all fit indices indicated adequate fit for the four-factor model within each ethnic group. As noted in the scaled score data analysis, Table 13 shows the magnitude of both $\underline{X}^2$ and $\underline{X}^2$/df varied with sample size. For the four-factor model, the $\underline{X}^2$/df index for the White group was of substantially less magnitude than for the one- or two-factor models. Given an acceptable $\underline{X}^2$/df index in the scaled score data analysis for the White group, it was concluded that the four-factor model also fit the data well, or at least fit best, for the White group. All subtest factor loadings were significant within each group.

Fit indices for the one- and two-factor models were not

Table 13

Within Groups Confirmatory Factor Analyses of Raw Data

One-Factor Model	Chi-Square	df	p	Chi-Square/df	TLI	AGFI
White	1664.56	65	<.001	25.61	.91	.78
Black	330.90	65	<.001	5.09	.93	.79
Hispanic	217.24	65	<.001	3.34	.94	.81
Two-Factor Model	**Chi-Square**	**df**	**p**	**Chi-Square/df**	**TLI**	**AGFI**
White	1067.08	64	<.001	16.67	.94	.86
Black	204.37	64	<.001	3.19	.96	.87
Hispanic	150.92	64	<.001	2.36	.97	.87
Four-Factor Model	**Chi-Square**	**df**	**p**	**Chi-Square/df**	**TLI**	**AGFI**
White	394.70	59	<.001	6.69	.98	.94
Black	118.30	59	<.001	2.01	.98	.92
Hispanic	100.20	59	<.001	1.70	.98	.91

consistently favorable. The TLI proved satisfactory for both models across all groups. X^2/df was satisfactory for the Black and Hispanic groups for the two-factor model. X^2/df was satisfactory for the Hispanic group when considering the one-factor model. The X^2/df index substantially exceeded 5.0 for the White group for both models. X^2/df barely exceeded 5.0 for the Black group one-factor model. The AGFI was less than .90 across all groups for both models. Given the mixed pattern of results for the one- and two- factor models, the four-factor model provided the most acceptable fit to the data.

General Findings: Simultaneous Analyses

Results from the simultaneous analyses indicated that the four-factor model exhibited structural and measurement invariance across the racial-ethnic groups. The three-factor model also proved invariant across groups. As mentioned above, the four-factor model was deemed most generalizable across groups. Results of the simultaneous analyses do not contradict this finding. Except for Object Assembly, the WISC-III subtests display adequate to ample subtest specificity across the age range (Sattler, 1992). In this study, all subtest factor loadings were significant for the four-factor solution during the within groups and simultaneous analyses. Therefore, the four-factor model is deemed the most generalizable and clinically useful model across White, Black, and Hispanic subgroups of the standardization sample. As in the within groups analysis, the one-, two-, and four-factor models are discussed in this section since they each use 13 subtests and are directly comparable. Results from analyses of the three-factor model are presented in the Appendices.

Scaled Score Data

Table 14 presents results of the simultaneous confirmatory factor analyses of scaled score data. Several model fit and incremental fit indices evaluated the factor models across groups. Models were evaluated as successive constraints were applied to each. The unconstrained model was evaluated first followed by constrained: (a) factor loadings; (b) factor loadings and error variances; (c) factor loadings, error variances, and factor variances; and (d) factor loadings, error variances, factor variances, and factor covariances (i.e., the "fully constrained" model). Factor covariances were not applicable to the one-factor model.

Results of the simultaneous analyses supported the fully constrained four-factor model. The model exhibited measurement and structural invariance across the White, Black, and Hispanic groups. Figures 5 and 6 illustrate the fully constrained four-factor model. Figure 5 displays the unstandardized solution. Figure 6 displays standardized factor subtest loadings and intercorrelations between factors. Hypothesis 1, the hypothesis that the

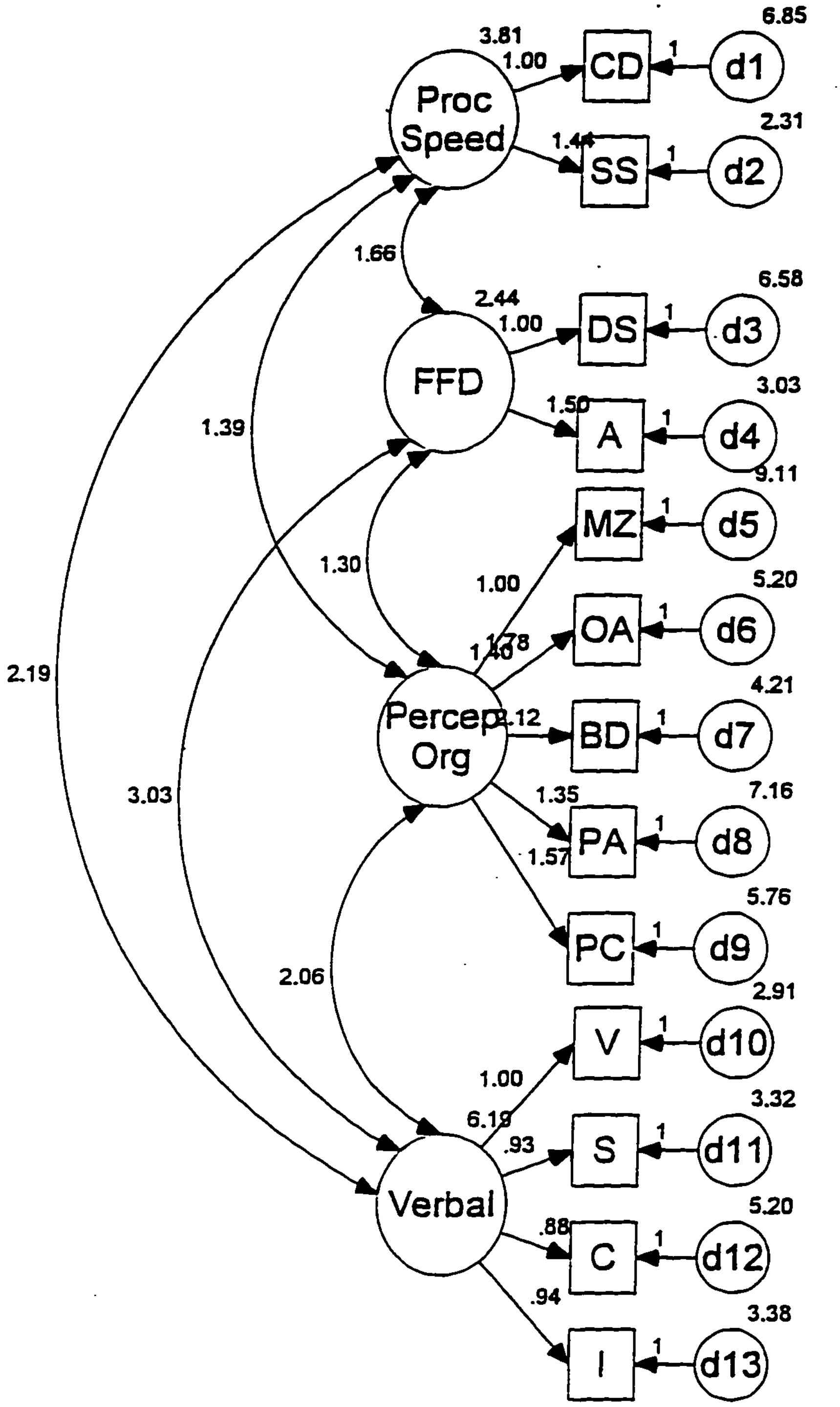
Table 14

Simultaneous Confirmatory Factor Analyses of Scaled Score Data Across White, Black, and Hispanic Subgroups*

Model	X^2	df	p	X^2/df	TLI	AGFI	ΔX^2	Δdf	p	$\Delta X^2/\Delta df$
One-Factor	1762.04	195	<.001	9.04	.80	.81	--	--	--	--
λ	1821.06	219	<.001	8.32	.82	.83	59.02	24	<.005	2.46
λ, δ	1881.86	245	<.001	7.68	.84	.84	60.80	26	<.005	2.34
λ, δ, Φ	1882.55	247	<.001	7.62	.84	.84	0.69	2	NS	0.35
Model	X^2	df	p	X^2/df	TLI	AGFI	ΔX^2	Δdf	p	$\Delta X^2/\Delta df$
Two-Factor	1022.97	192	<.001	5.33	.89	.90	--	--	--	--
λ	1067.45	214	<.001	4.99	.90	.90	44.48	22	<.005	2.02
λ, δ	1125.84	240	<.001	4.69	.91	.91	58.39	26	<.005	2.25
λ, δ, Φ	1142.63	244	<.001	4.68	.91	.91	16.79	4	<.005	4.20
$\lambda, \delta, \Phi, \Phi_c$	1146.30	246	<.001	4.66	.91	.91	3.67	2	NS	1.84
Model	X^2	df	p	X^2/df	TLI	AGFI	ΔX^2	Δdf	p	$\Delta X^2/\Delta df$
4-Factor	422.98	177	<.001	2.39	.97	.95	--	--	--	--
λ	468.14	195	<.001	2.40	.97	.95	45.16	18	<.005	2.51
λ, δ	524.83	221	<.001	2.37	.97	.95	55.69	26	<.005	2.14
λ, δ, Φ	556.64	229	<.001	2.43	.97	.95	31.81	8	<.005	3.98
$\lambda, \delta, \Phi, \Phi_c$	574.30	241	<.001	2.38	.97	.96	17.66	12	NS	1.47

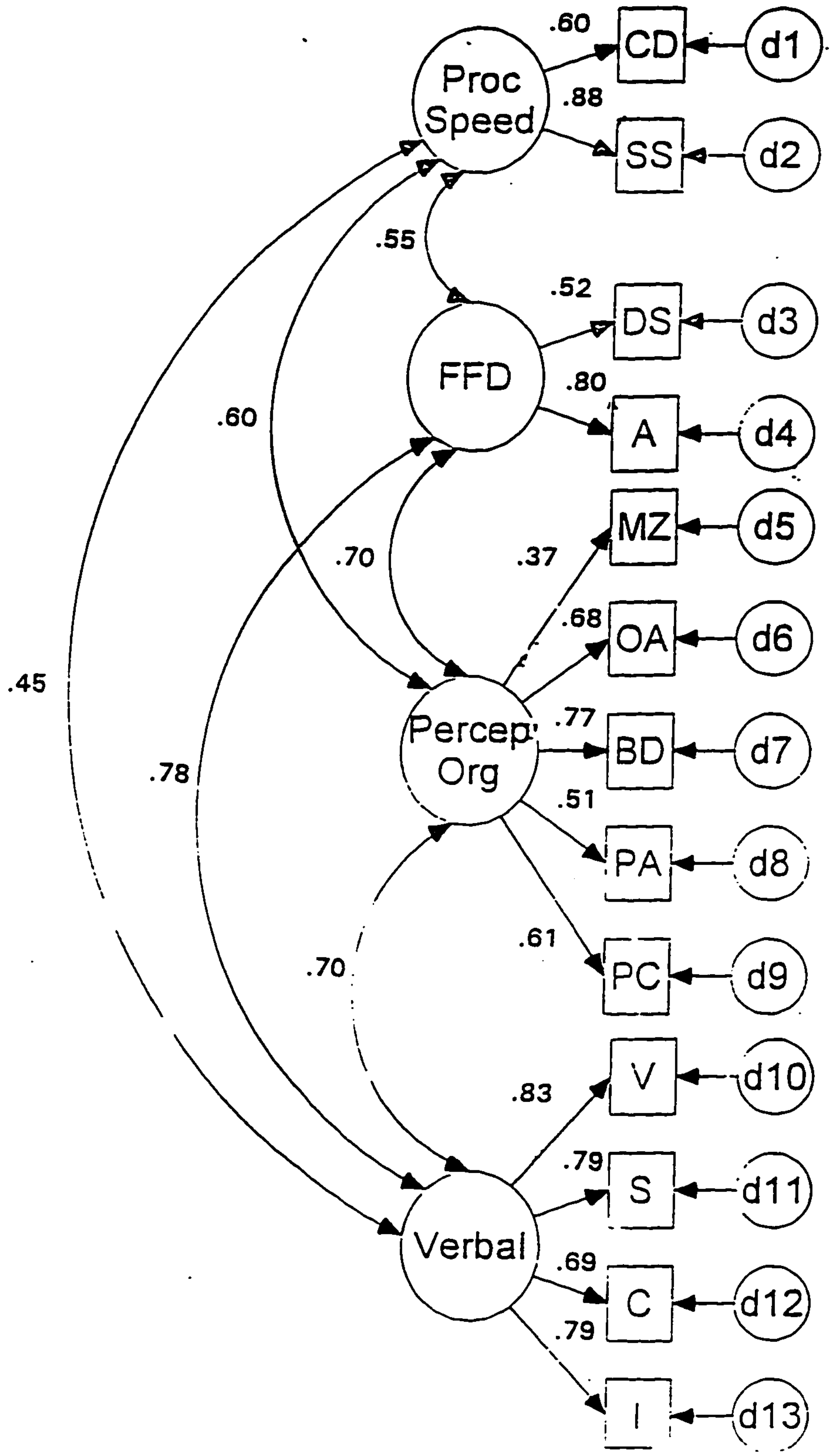
Note. λ = constrained factor loadings; δ = constrained error variances; Φ = constrained factor variances; Φ_c = constrained factor covariances.

Figure 5 Fully Constrained Four-Factor Model: Scaled Score Data



Chi-Square = 574.30 Chi-Square/df = 2.38
df = 241 TLI = .97
p < .001 AGFI = .96

Figure 6 Standardized Factor Loadings For Four-Factor Model: Scaled Score Data



unconstrained two-factor model would provide for the best data fit across subgroups, was not supported.

Assessment for invariance began with inspection of fit indices for the unconstrained model. This initial step determined how many factors were needed to explain observed subtest variance-covariance matrices across all groups. Table 14 shows that all model fit indices (except AGFI for the two-factor model) for the unconstrained one- and two-factor models were not acceptable. Except for the summative $\underline{X}^2$, fit indices for the unconstrained four-factor model were favorable. This means that more than two factors are necessary to explain the data across groups. This result was not surprising given that the four-factor model fit scaled and raw data variance-covariance matrices best during within groups analysis.

The constrained four-factor model was evaluated by model fit and incremental fit indices $\Delta\underline{X}^2$, $\Delta\underline{X}^2/\Delta$df. These indices are discussed on pp. 84-87. Incremental fit indices evaluated whether significant differences existed between more constrained models and lesser constrained models.

As seen in Table 14, when factor loadings were constrained to be equal across groups acceptable $\underline{X}^2$/df, TLI, and AGFI were obtained for the four-factor model. The incremental fit index $\Delta\underline{X}^2$ was significant but the $\Delta\underline{X}^2/\Delta$df was acceptable (i.e., less than 5.0). Like $\underline{X}^2$, $\Delta\underline{X}^2$ is influenced by sample size. Given that all other model fit and incremental fit indices were acceptable, it was concluded that the four-factor model with constrained factor loadings was a good fit across groups. This suggests that the same pattern of factor loadings exists for each group on each of the four factors.

Gradually increased constraints on the four-factor model produced a similar pattern of results. $\underline{X}^2$ and $\Delta\underline{X}^2$ remained significant for the four-factor model with constrained error variances. However, $\underline{X}^2$/df, TLI, AGFI, and $\Delta\underline{X}^2/\Delta$df indicated acceptable fit. This suggests that subtests in the four-factor model are measured with the same reliability across groups.

Favorable results for the four-factor model were evidenced when constrained factor variances and factor covariances were added in succession. The pattern of model and incremental fit indices shown in Table 14 suggest that White, Black, and Hispanic groups exhibit the same spread of factor scores and the same factor covariances. Except for $\underline{X}^2$, all model and incremental fit indices were acceptable for the fully constrained four-factor model. Measurement invariance across groups was concluded given the equality of factor loadings and subtest reliabilities. Structural invariance was concluded given the equality of number of factors, factor variances, and factor covariances across groups.

Raw Data

Table 15 presents results of the simultaneous analysis of raw data. Again, measurement and structural invariance was evidenced by four-factor model. The fully constrained four-factor model is pictured in Figures 7 and 8. The unstandardized solution is depicted in Figure 7 and the standardized factor loadings and factor intercorrelations are seen in Figure 8. The unconstrained one- and two-factor models did not fit the data well across groups. This pattern of results matches those found in the simultaneous analysis of scaled score data. Taken together, analyses of scaled and raw data do not support the hypotheses of this study.

Model and incremental fit indices for the four-factor model displayed the same pattern as in the simultaneous analysis of scaled score data. $\underline{X}^2$ was significant at all levels of constraint. $\Delta\underline{X}^2$ was significant at all levels of constraint except for the fully constrained model. $\underline{X}^2$/df, TLI, AGFI, and $\Delta\underline{X}^2/\Delta$df indicated good fit at all levels of constraint.

Subtest Factor Loadings

Table 16 shows unstandardized subtest factor loadings and their relative rank order on each factor of the fully constrained four-factor model. Readers more familiar with standardized factor loadings and factor correlations should refer to Appendices I and J for estimated values. Analyses

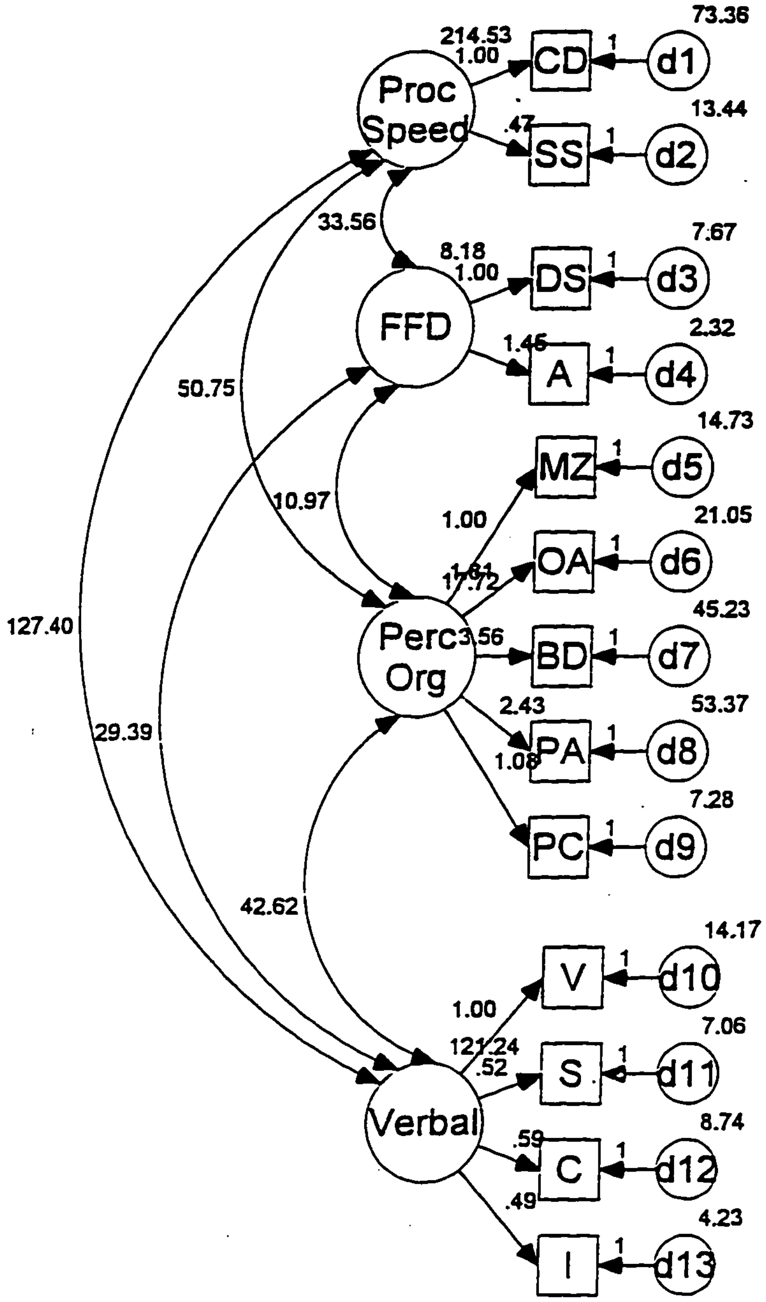
Table 15

Simultaneous Confirmatory Factor Analyses of Raw Data Across White, Black, and Hispanic Subgroups

Model	X^2	df	p	X^2/df	TLI	AGFI	ΔX^2	Δdf	p	$\Delta X^2/\Delta$df
One-Factor	2212.54	195	<.001	11.35	.92	.78	--	--	--	--
λ	2293.68	219	<.001	10.47	.92	.80	81.14	24	<.005	3.38
λ,δ	2343.03	245	<.001	9.56	.93	.82	49.35	26	<.005	1.90
λ,δ,Φ	2348.56	247	<.001	9.51	.93	.82	5.53	2	NS	2.76
Model	X^2	df	p	X^2/df	TLI	AGFI	ΔX^2	Δdf	p	$\Delta X^2/\Delta$df
Two-Factor	1422.29	192	<.001	7.42	.95	.86	--	--	--	--
λ	1479.83	214	<.001	6.92	.95	.87	57.54	22	<.005	2.62
λ,δ	1530.03	240	<.001	6.38	.96	.88	50.20	26	<.005	1.93
λ,δ,Φ	1554.33	244	<.001	6.37	.96	.88	24.30	4	<.005	6.08
$\lambda,\delta,\Phi,\Phi_c$	1556.41	246	<.001	6.33	.96	.88	2.08	2	NS	1.04
Model	X^2	df	p	X^2/df	TLI	AGFI	ΔX^2	Δdf	p	$\Delta X^2/\Delta$df
4-Factor	613.34	177	<.001	3.46	.98	.93	--	--	--	--
λ	669.30	195	<.001	3.43	.98	.93	55.96	18	<.005	3.11
λ,δ	730.08	221	<.001	3.30	.98	.94	60.78	26	<.005	2.34
λ,δ,Φ	761.99	229	<.001	3.33	.98	.94	31.91	8	<.005	3.99
$\lambda,\delta,\Phi,\Phi_c$	781.78	241	<.001	3.24	.98	.94	19.79	12	NS	1.65

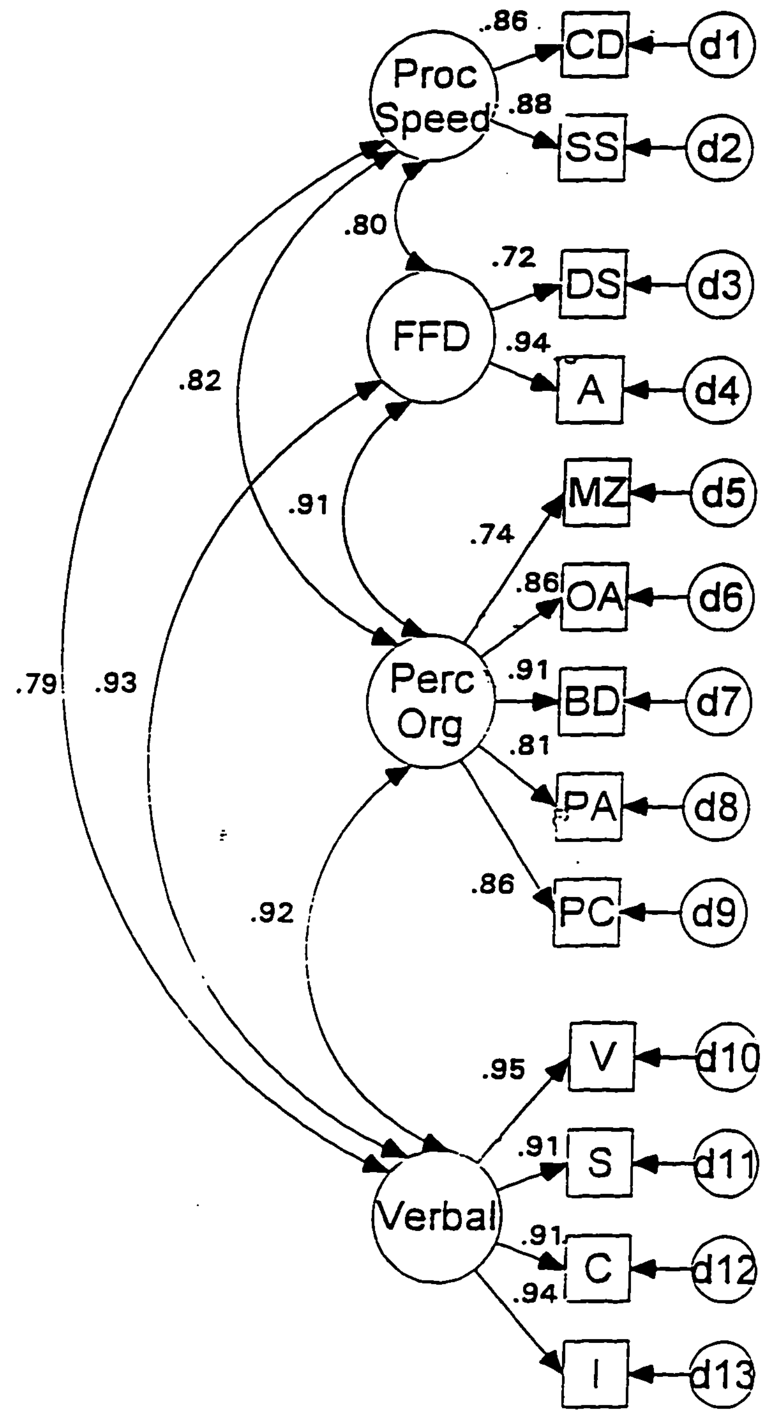
Note. λ = constrained factor loadings; δ = constrained error variances; Φ = constrained factor variances; Φ_c = constrained factor covariances.

Figure 7 Fully Constrained Four-Factor Model: Raw Score Data



Chi-Square = 781.78 Chi-Square/df = 3.24
df = 241 TLI = .98
p < .001 AGFI = .94

Figure 8 Standardized Factor Loadings For Four-Factor Model: Raw Score Data



of rank order allowed for further comparisons between results from the raw and scaled score data analyses. Minor differences in relative rank orders were observed on the Verbal and Perceptual Organization factors. The rank order of Coding and Symbol Search subtests on the Processing Speed factor depended upon the dataset analyzed. All subtest factor loadings on Table 16 were statistically significant.

Some consistencies in the relative rank orders were evident. Subtest factor loadings displayed 50 percent agreement in rank order across datasets on the Verbal factor. On this factor, Vocabulary emerged with the highest factor loading for both raw and scaled score data. On the Perceptual Organization factor, 40 percent agreement in subtest rank order was found. Block Design had the highest factor loading on the Perceptual Organization factor across datasets. Mazes consistently had the smallest factor loading on this factor. Arithmetic consistently had the higher factor loading on the Freedom From Distractibility factor.

Since minor differences in relative rank order of subtest factor loadings were detected between datasets for the fully constrained four-factor models. Subtest rank orders were inspected within each racial-ethnic group. Subtest factor loadings from the within groups analysis were analyzed. Again, Hypothesis 3 was not supported. In fact the opposite was observed. More differences in relative

Table 16

Subtest Unstandardized Factor Loadings for Fully Constrained Four-Factor Model: Raw Data vs. Scaled Score Data*

Factor	Raw Data	Scaled Score Data
Verbal		
Vocabulary	1.00 (1)	1.00 (1)
Similarities	0.52 (3)	0.93 (3)
Comprehension	0.59 (2)	0.88 (4)
Information	0.49 (4)	0.94 (2)
Perceptual Organization		
Mazes	1.00 (5)	1.00 (5)
Object Assembly	1.81 (3)	1.78 (2)
Block Design	3.56 (1)	2.12 (1)
Picture Arrangement	2.43 (2)	1.35 (4)
Picture Completion	1.08 (4)	1.57 (3)
Freedom From Distractibility		
Digit Span	1.00 (2)	1.00 (2)
Arithmetic	1.46 (1)	1.50 (1)
Processing Speed		
Coding	1.00 (1)	1.00 (2)
Symbol Search	0.47 (2)	1.44 (1)

*Note. Rank order of the subtest is in parentheses for each factor. The first subtest listed under each factor had its factor loading set to 1.00. This provided a metric on which all other factor loadings for a given factor were based.

rank orders of subtest factor loadings were found in the scaled score data analysis. This finding was unexpected and is addressed later in the discussion.

Table 17 presents subtest factor loadings for the White group four-factor model. These obtained factor loading patterns are identical to the patterns obtained in the simultaneous analysis across factors and datasets. The White group's large sample size (N=1543) relative to the Black (N=338) and Hispanic (N=242) groups probably contributed to this finding.

Table 18 presents factor loadings for the Black group. Similarities and differences were noted as compared to the White group's and fully constrained four-factor model. Within the Black group, the factor loading pattern differed between the raw and scaled score datasets (50 percent agreement on the Verbal factor, 60 percent on Perceptual Organization, 100 percent on Freedom From Distractibility, no agreement on Processing Speed). The pattern of obtained factor loadings from the raw data analysis is identical to the White group. The pattern seen in the Black group scaled score data is different from the White group for the Verbal (25 percent agreement) and Perceptual Organization factors (60 percent agreement). The key similarities within these factors are that Vocabulary has the highest subtest loading for the Verbal factor and Block Design is the best measure of Perceptual Organization. Only a minor difference was

Table 17

Subtest Unstandardized Factor Loadings for White Sample Four-Factor Model: Raw Data vs. Scaled Score Data*

Factor	Raw Data	Scaled Score Data
Verbal		
Vocabulary	1.00 (1)	1.00 (1)
Similarities	0.52 (3)	0.95 (3)
Comprehension	0.58 (2)	0.90 (4)
Information	0.49 (4)	0.97 (2)
Perceptual Organization		
Mazes	1.00 (5)	1.00 (5)
Object Assembly	1.84 (3)	1.93 (2)
Block Design	3.71 (1)	2.26 (1)
Picture Arrangement	2.54 (2)	1.39 (4)
Picture Completion	1.10 (4)	1.56 (3)
Freedom From Distractibility		
Digit Span	1.00 (2)	1.00 (2)
Arithmetic	1.45 (1)	1.59 (1)
Processing Speed		
Coding	1.00 (1)	1.00 (2)
Symbol Search	0.46 (2)	1.46 (1)

*Note. Rank order of the subtest is in parentheses for each factor. The first subtest listed under each factor had its factor loading set to 1.00. This provided a metric on which all other factor loadings for a given factor were based.

Table 18

Subtest Unstandardized Factor Loadings for Black Sample Four-Factor Model: Raw Data vs. Scaled Score Data

Factor	Raw Data	Scaled Score Data
Verbal		
Vocabulary	1.00 (1)	1.00 (1)
Similarities	0.50 (3)	0.81 (4)
Comprehension	0.62 (2)	0.89 (2)
Information	0.45 (4)	0.83 (3)
Perceptual Organization		
Mazes	1.00 (5)	1.00 (5)
Object Assembly	1.63 (3)	1.34 (3)
Block Design	3.02 (1)	1.59 (1)
Picture Arrangement	1.96 (2)	1.08 (4)
Picture Completion	1.01 (4)	1.41 (2)
Freedom From Distractibility		
Digit Span	1.00 (2)	1.00 (2)
Arithmetic	1.37 (1)	1.10 (1)
Processing Speed		
Coding	1.00 (1)	1.00 (2)
Symbol Search	0.46 (2)	1.21 (1)

*Note. Rank order of the subtest is in parentheses for each factor. The first subtest listed under each factor had its factor loading set to 1.00. This provided a metric on which all other factor loadings for a given factor were based.

observed in the Perceptual Organization factor between the Black and White groups. Picture Completion has a higher loading than Object Assembly in the Black group: the reverse is true for the White group. Freedom From Distractibility and Processing Speed factor loading patterns were identical.

Table 19 presents subtest factor loadings for the Hispanic group. Within the Hispanic group, factor loading patterns differed between raw and standardized datasets (25 percent agreement on the Verbal factor, 60 percent agreement on Perceptual Organization, 100 percent on Freedom From Distractibility, and 0 percent on Processing Speed). The pattern of obtained factor loadings for raw data is identical to that seen in the White and Black group raw data analyses. The Perceptual Organization factor loading pattern for scaled score data is identical to the Black group's pattern on this factor (also 60 percent agreement with the White group). Freedom From Distractibility and Processing Speed factor loading patterns are identical to the patterns in the Black and White groups. The Hispanic Verbal factor differs from the Verbal factors in the Black and White groups (50 percent agreement with both groups). However, Vocabulary is also the best measure of the Verbal factor in the Hispanic group as it was for the Black and White groups.

Table 19

Subtest Unstandardized Factor Loadings for Hispanic Sample Four-Factor Model: Raw Data vs. Scaled Score Data

Factor	Raw Data	Scaled Score Data
Verbal		
Vocabulary	1.00 (1)	1.00 (1)
Similarities	0.57 (3)	0.99 (2)
Comprehension	0.64 (2)	0.73 (4)
Information	0.53 (4)	0.86 (3)
Perceptual Organization		
Mazes	1.00 (5)	1.00 (5)
Object Assembly	1.86 (3)	1.75 (3)
Block Design	3.46 (1)	2.35 (1)
Picture Arrangement	2.46 (2)	1.72 (4)
Picture Completion	1.09 (4)	2.07 (2)
Freedom From Distractibility		
Digit Span	1.00 (2)	1.00 (2)
Arithmetic	1.55 (1)	1.59 (1)
Processing Speed		
Coding	1.00 (1)	1.00 (2)
Symbol Search	0.54 (2)	1.79 (1)

*Note. Rank order of the subtest is in parentheses for each factor. The first subtest listed under each factor had its factor loading set to 1.00. This provided a metric on which all other factor loadings for a given factor were based.

<u>Summary of Results</u>

The four-factor model demonstrated structural and measurement invariance across White, Black, and Hispanic subgroups of the standardization sample. This result was evident for both scaled and raw score data. Identical Verbal Comprehension, Perceptual Organization, Freedom From Distractibility, and Processing Speed factors were found across groups when raw score data were analyzed. Unique Verbal Comprehension factors were observed for each group when scaled score data was analyzed. Black and Hispanic groups displayed identical Perceptual Organization factors. The White group's Perceptual Organization factor differed only slightly from both the Black and Hispanic groups. Freedom From Distractibility and Processing Speed were identical across groups.

Chapter IV

DISCUSSION

Evaluation of the Hypotheses

The present study demonstrates the structural and measurement invariance of the four-factor model across White, Black, and Hispanic subgroups of the standardization sample. The four-factor model exhibits invariance across scaled score and raw score datasets. Relative rank orders of subtest factor loadings are consistent across groups for the raw score dataset. Some differences in rank order are evident for the scaled score dataset on the Verbal Comprehension and Perceptual Organization factors. These results do not support the hypotheses of this study.

The hypothesis that a two-factor model would best fit each group's data, Hypothesis 1, was not supported. The verbal-performance model demonstrated inadequate fit across groups and datasets. Simultaneous confirmatory factor analysis indicated that one- and two-factor models were insufficient to explain White, Black, and Hispanic subgroup's data. Within group analyses indicated good fit for the three- and four-factor models. However, the four-factor model demonstrated the best fit for all groups. Simultaneous analyses supported this result.

Hypothesis 2, that factors would be measured by different scaling units (i.e., factor loadings) was also not supported. The pattern of subtest factor loadings conform to the four-factor model for all groups. Verbal Comprehension consists of Information, Similarities, Vocabulary, and Comprehension. Perceptual Organization consists of Picture Arrangement, Picture Completion, Block Design, Object Assembly, and Mazes. Freedom From Distractibility consists of Arithmetic and Digit Span. Processing Speed consists of Coding and Symbol Search. This is the case regardless of analyzing raw scores or scaled scores.

Hypothesis 2 suggested a lack of measurement invariance for the factor model. However, results demonstrate complete measurement invariance across racial-ethnic groups. The four factors are measured in the same scaling units with the same degree of reliability. The 13 WISC-III subtests tap the same true score variance for the three groups. This means that WISC-III test scores may be interpreted the same way across the racial-ethnic subgroups of this study.

Structural invariance of the four-factor model was also found. Factor variances and covariances are equivalent across the three groups. This means that factor score variances and the interrelationships between factors are equal across groups. This has implications not only for psychological assessment practices but for Wechsler's conceptualization of intelligence which is discussed below.

Hypothesis 3 was also not supported. This concerned the expectation that raw score data analysis would reveal idiosyncratic response patterns within each of the three racial-ethnic groups. It was expected that differences in factor loading patterns would be greater when raw data were analyzed. However, factor loadings patterns are equivalent across groups when one analyzes raw data. This includes the relative rank orders of factor loading magnitude for subtests.

Differences in rank order of subtest factor loadings is evident when scaled score data are analyzed. It was reasoned above that when raw scores are converted to scaled scores during standardization, that unique response patterns of the smaller minority samples would be masked by the larger White group. It was expected that the groups would appear more homogeneous with respect to ability profiles with analysis of scaled scores. This is not the case: the groups appear more different when scaled scores are analyzed.

In the sections that follow, theoretical and methodological issues attempt to account for the present results. Theoretical issues involve Wechsler's conceptualization of intelligence. Multicultural issues are also discussed with respect to acculturation. Methodological issues concern characteristics of the present sample including issues of defining multicultural terminology and socioeconomic status (SES). The use of confirmatory factor analysis in comparison to exploratory factor analysis in determining the factor structure of a test is also discussed.

Theoretical Issues

Wechsler's Conceptualization of Intelligence

Wechsler's original verbal-performance dichotomy remains largely intact as evidenced by the Verbal Comprehension and Perceptual Organization factors. Freedom From Distractibility and Processing Speed add to the clinical utility of the WISC-III. Whether Freedom From Distractibility and Processing Speed account for Wechsler's nonintellective factors is debatable. They do not seem to directly measure motivation or emotional factors that affect test performance. However, performance on Freedom From Distractibility and Processing Speed (attention to task and speed of perceptual processing, respectively; Sattler, 1988; Sattler, 1992) may allow a clinician to infer motivational factors.

This study supports the cross racial-ethnic validity of Wechsler's conceptualization of intelligence. Previous work with the WISC-R generally supported verbal and performance factors across racial-ethnic subgroups. However, exploratory factor analyses suggested different scaling units for the factors depending on racial-ethnic group membership. This study, using a large nationally representative sample of White, Black, and Hispanic children, provides evidence for the equality of factor composition across groups. The addition of Freedom From Distractibility and Processing Speed indicate that Wechsler's original conceptualization of a verbal-performance dichotomy is incomplete. More than two factors are needed to account for intellectual performance, at least on the WISC-III.

The Multicultural Literature

The multicultural literature review above provides a cogent argument for the expectation that structural and measurement invariance would not hold across the groups studied. Far too many variables are said to account for cognitive skills development in children. These nature of these variables differ depending on the cultural setting in which the child is raised. Minority group's distance from majority group values (Grubb & Dozier, 1989; Hinkle, 1994), learning styles (Schiele, 1991), the test-taking context (Miller-Jones, 1989; Rodriguez, 1992), and different learning histories across cultural groups (Geisinger, 1992) all could form different patterns of ability expression across racial-ethnic subgroups. Recently, Steele (1997) discussed stereotypes and their often detrimental influence on minority groups performance on standardized tests.

The present study suggests that White, Black, and Hispanic subgroups are more similar than they are different with respect to intellectual expression. Sensitivity to the issues above prompted the WISC-III development team to review potentially biased test items. Wechsler (1991) describes procedures to safeguard against item bias during WISC-III development. Suzuki, Vraniak, & Kugler (1996) also discuss the process. Subjective analyses and empirical pilot trials with revised items from the WISC-R allowed for a fairer test across racial-ethnic groups. Persons and subject matter are intended to appeal to a multicultural society. Helms (1992) states that the WISC-III's diverse cultural content is one way to facilitate equivalent assessment of Black children. It is argued here that the same may be said to extend to Hispanic children. The WISC-III appears to be a product of state-of-the-art psychometric development that successfully measures its constructs the same way across different racial-ethnic groups.

Fairness in predictive validity of the WISC-III for academic achievement was demonstrated by Weiss, Prifitera, & Roid (1993). No differences were found in predicting achievement from WISC-III performance across nationally representative White, Black, and Hispanic groups. Similarity of constructs and in prediction across groups expands the clinical utility of the WISC-III. Whether test development procedures aimed to equate groups on psychological constructs blurs unique aspects between groups remains an empirical question.

The present study provides evidence for the conceptual equivalence of the WISC-III across the racial-ethnic groups studied. Conceptual equivalence means that a test's constructs has the same meaning across different cultural groups (Okazaki & Sue, 1995). This means that Verbal Comprehension, Perceptual Organization, Freedom From Distractibility, and Processing Speed appear to have the same meaning across the White, Black, and Hispanic groups. However, before concluding that the four-factor model is conceptually equivalent across groups, one must acknowledge that similar factor structure may be due to acculturation factors not accounted for by this study. It is possible that the present results reflect the organization of knowledge of mainstream culture rather than a true organization of intelligence.

Berry et al. (1992) describes one type of acculturation known as integration. This is acculturation whereby different racial-ethnic groups interact and share common societal endeavors yet attempt to retain their distinctiveness. Inclusion of test content that appeals to a pluralistic society may tap into this acculturation process. Helms (1992) discusses geographic proximity of different social groups, including intermixing, and the impact this may have on psychometric outcomes. Integration can obviously result in idea exchange, but it may also result in nonindependence of groups, which has psychometric consequences. Rather than getting pure qualitative assessments of intelligence structures, we may obtain patterns reflective of varying degrees of familiarity with mainstream culture.

Sternberg (1990) discusses models of intelligence assessment regarding emic/etic measures. One model posits that the same test can be used cross-culturally to assess the constructs that are said to exist universally. This would lead one to conclude that assessment may proceed without regard given to a child's cultural background. This study provides evidence that one test may be used across different racial-ethnic groups. However, it does not necessarily reflect the model outlined by Sternberg (1990).

Different cultures were not assessed in this study. Teo & Febbraro (1997) state that each racial-ethnic group in society has significant within group variability in cultural practices. Given the sampling procedure discussed in Wechsler (1991), it may be assumed that heterogeneity existed within each racial-ethnic subgroup. For example, within the Hispanic racial-ethnic subgroup it is likely that Puerto Rican, Central American, and Cuban, children were included in the standardization sample (to mention just a few). Within these subgroups of Hispanics exist families who express different cultural values and execute different cultural practices. This significant variability was masked in this study by the lumping of all children into three large racial-ethnic categories. Therefore, individualized assessment should still give large regard to cultural background.

Methodological Issues

Within Group Heterogeneity

The present study compared three racial-ethnic groups. As noted above, each of these groups are heterogeneous with respect to ethnicity and culture. In addition, each group represents a conglomerate of SES levels. Whereas individuals of the same ethnicity may engage in different cultural practices (e.g., Teo & Febbraro, 1997) persons of different SES levels may have differential access to educational opportunity. In either case, different learning styles and ability development may occur. Although Wechsler (1991) states that no relationship exists between SES and WISC-R factor structure, this remains an empirical question for the WISC-III. Given the disproportion of minority children in lower SES brackets, it would be interesting to study the factor structure of the WISC-III in large samples broken down by race-ethnicity and SES.

This may continue to be a difficult endeavor as long as terms such as "race," "ethnicity," and "culture" continue to be poorly defined in psychological research. This study is no exception. Okazaki & Sue (1995) discuss how ethnicity is a demographic variable that is often used as a proxy for culture. The three racial-ethnic groups distinguished here is consistent with previous work and was indeed used to infer different cultural practices at a macro level. Until these terms are more carefully defined, and subgroups within larger racial-ethnic groups are isolated (thereby breaking down the heterogeneity in studies like this one) the effect that cultural practices have on intellectual expression remains an empirical question. As this study suggests, when groups are studied at a macro level, no differences in intellectual expression exist. Therefore, the next step calls for deeper probing.

Specific age levels within the 6-16 year span were not examined in this study. Wechsler (1991) did not find differences in factor structure across the 6-7, 8-10, 11-13, and 14-16 year age spans. However, it is possible that differences in factor structure may exist within and between racial-ethnic groups at various ages. Helms (1992) refers to acculturation as a dynamic process. It would be interesting to see if factor structure becomes more similar between groups over time. In order to run analyses broken down by age and race-ethnicity much larger standardization samples are required.

Statistical Issues

Differences in subtest rank orders exist between raw and scaled score data analyses within each racial-ethnic group. However, results from the White group's raw score and scaled score within groups analyses match results of the respective simultaneous analyses. For Black and Hispanic groups, their respective within groups analyses differed from the simultaneous analyses. It appears that alterations in the multivariate distributions of each group is influenced greatly by the relatively large White subgroup when raw scores were converted to scaled scores. Inspection of raw score results indicate equivalence of subtest rank order across groups for within group and simultaneous analyses.

It appears that some information about each subgroup's factor loading profile may be lost when raw score datasets are pooled to create a single set of standardized scores. Structural and measurement properties and the relative rank order of subtest factor loadings are invariant when raw data are analyzed. Structural and measurement invariance holds when scaled scores are used. However, the qualitative emphasis of each factor may appear different as evidenced by differences in rank order of subtest factor loadings. For example, Vocabulary is consistently the best measure of Verbal Comprehension. However, the second best measure differs across groups: Information for the White group; Comprehension for the Black group; and Similarities for the Hispanic group. This is not the case when raw data are analyzed.

In the context of the present study's findings, this point may appear minor. However, raw score data from a nationally representative sample does not ensure similar multivariate distributions across subsamples (i.e., similar means and standard deviations). Therefore, when data from the subsamples are lumped together to create scaled scores, the data transformation may alter subtest covariances differentially across groups. This is probably more true for groups of smaller sample size. Therefore, one may expect factor profiles that are not homogeneous when scaled scores are analyzed. In this study, the multivariate distributions were not available from the test publishers.

This study made use of advances in factor analysis. Confirmatory factor analysis has three major advantages over exploratory factor analysis which is pervasive in the Wechsler psychometric literature. The first advantage is that confirmatory factor analysis allowed the testing of an a priori specified model. The test allowed a determination if the model proposed left significant variance unaccounted for (Floyd & Widaman, 1995). This is in contrast to what exploratory analyses accomplish: the search for significant variance accounted for by factors. While exploratory factor analysis allows for a determination of significant variance accounted for by factors, it does not directly address variance unaccounted for. In this study, one can safely conclude that Verbal Comprehension, Perceptual Organization, Freedom From Distractibility, and Processing Speed account for significant variance and do not leave much variance unaccounted for in the subtest variance-covariance matrices of each racial-ethnic subgroup.

The second major advantage to confirmatory factor analysis, evidenced in this study, is the testing of a priori models. Judgments as to how may factors to retain was not an issue. Sattler (1992) rejected Freedom From Distractibility as a factor due to an arbitrary eigenvalue criterion. However, this study's confirmatory approach did not leave the inclusion of any factor to judgment. A factor model is accepted or rejected in total. In this case, the four-factor model is accepted.

A final advantage of confirmatory factor analysis is its ability to assess for measurement invariance. Previous exploratory work with the WISC-R cited equivalence of factor structure across racial-ethnic subgroups. However, that body of literature could not assess for measurement invariance with exploratory factor analysis. This study allows for a determination, not only of the equivalence of factor structure, but also the equivalence of scaling units and reliability.

Application of the Present Study's Findings

Utility of the Four-Factor Model

The present study extends the findings of Wechsler (1991) and Roid et al. (1993) in support of the four-factor model. Unlike the three-factor model proposed by Sattler (1992), the four-factor model employs all 13 WISC-III subtests. This enhances the clinical utility of the WISC-III. The confirmatory factor analyses of this study supports the existence of the controversial Freedom From Distractibility factor. This study also supports the existence of the new Processing Speed factor. In addition, the study supports the inclusion of two subtests recently described as problematic: Mazes and Picture Arrangement.

The Mazes subtest does not contribute to the computation of the Perceptual Organization index score. Despite satisfactory psychometric properties Wechsler (1991) deleted this subtest in the computation of index scores. Wechsler (1991) states that Mazes is often the lowest loading subtest on the Perceptual Organization scale. It is also frequently deleted from empirical analyses since it is a supplementary subtest. Mazes consistently had the lowest subtest factor loading on the Perceptual Organization factor in this study as well. However, its loading was nonetheless significant for all racial-ethnic groups in this study. This evidence argues for its inclusion in the Perceptual Organization index score.

Kamphaus (1993) cites Picture Arrangement as a difficult subtest to interpret. He discusses the Wechsler (1991) exploratory analyses in which the subtest loaded significantly on both the Perceptual Organization factor and the Verbal Comprehension factor. This pattern held for both the two-factor and four-factor solutions presented in the manual. This is comparable to the Reynolds & Ford (1994) exploratory factor analysis calling for a three-factor solution. Again, Picture Arrangement had its highest loading on Perceptual Organization, but also had a significant secondary loading on Verbal Comprehension.

When Kamphaus et al. (1994) performed their confirmatory factor analyses on data from Wechsler (1991), Picture Arrangement loaded significantly on the Perceptual Organization factor. The present study also finds Picture Arrangement to load significantly on Perceptual Organization. This is the case across the White, Black, and Hispanic groups. Methodological differences between exploratory and confirmatory factoring techniques may impact the interpretation of Picture Arrangement. Another key difference is that Wechsler (1991) and Reynolds & Ford (1994) used varimax rotation which results in uncorrelated factors. In this study and in Kamphaus et al. (1994), factor models allowed for correlated factors.

In clinical practice, the four-factor interpretation may be applied to White, Black, and Hispanic subgroups. However, it is recommended that a child's cultural background be assessed. Even though the WISC-III factors are measured equally across racial-ethnic subgroups of the standardization sample, the study did not examine factors which may affect quantitative performance on the test.

Conclusion

This study demonstrates the structural and measurement invariance of the WISC-III across White, Black, and Hispanic subgroups of the standardization sample. Confirmatory factor analysis shows the Wechsler (1991) four-factor model to be the best fitting model among three other competing models. Verbal Comprehension, Perceptual Organization, Freedom From Distractibility, and Processing Speed account well for the scaled score and raw score variance-covariance matrices of each racial-ethnic group.

Results of this study do not support the original hypotheses. The multicultural literature reviewed earlier led to the expectation that the organization of cognitive abilities in children of different racial-ethnic groups would differ. This is not the case for the abilities measured by the WISC-III. Several reasons are offered to account for the findings.

The daily integration of different racial-ethnic groups may account for the similarities observed in this study. The age of rapid information exchange facilitates the intermixing of people from different backgrounds. Acculturation becomes easier in the form of integration rather than separation or marginalization. People from different cultural backgrounds may become more familiar with each other's practices and values. As a result, test developers become aware of test content more suitable for a multicultural (or pluralistic) society.

It must be noted that awareness of a child's cultural background remains paramount during individualized clinical assessment. The review noted above indicates the problems with empirical work using poorly defined terms such as "race," "ethnicity," and "culture." Race-ethnicity is often used as a proxy for culture, which is a mistake. Once empirical efforts are better able to discriminate between cultural practices we may be in better position to study less heterogeneous groups. Until then, research will probably continue to analyze data at a macro level from "racial-ethnic" groups.

It is possible that subgroups of particular age, SES, and cultural backgrounds differ in their intellectual makeup. Or at least they may differ in intellectual expression. Again, this study was conducted at a macro level across the 6-16 year age span. Clinical practice must continue to take a child's background into account. This study's findings should not be misinterpreted as to disregard the influence a child's background will have on his or her expression of intelligence.

Future research must attempt to clarify terms used to distinguish groups in multicultural research. Only then can work be done to identify psychological characteristics of smaller subgroups in society. Until this is done, a great many may not be provided with adequate psychological services. The present study is a step in that direction.

Future research should also attempt validity studies within racial-ethnic groups during test development. As this study shows, results may differ to some extent depending on the type of data analyzed. Although structural and measurement invariance held across raw and scaled score datasets, differences were noted in rank order of factor loading magnitude. Researchers should be aware that standardizing raw scores may differentially alter the intercorrelations of observed variables within subgroups of a larger sample.

References

Allen, S.R. & Thorndike, R.M. (1995). Stability of the WPPSI-R and WISC-III factor structure using cross-validation of covariance structure models. Journal of Psychoeducational Assessment, 13, 3-20.

Arbuckle, J.L. (1997). Amos Users' Guide Version 3.6. Chicago: SmallWaters Corporation.

Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88(3), 588-606.

Berry, J.W., Poortinga, Y.H., Segall, M.H., & Dasen, P.R. (1992). Cross-cultural psychology: Research and applications. Cambridge University Press: New York.

Blaha, J., Wallbrown, F.H., & Wherry, R. (1974). Hierarchical factor structure of the Wechsler Intelligence Scale for Children. Psychological Reports, 35, 771-778.

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

Butler-Omololu, C., Doster, J.A., & Lahey, B. (1984). Some implications for intelligence test construction and administration with children from different racial groups. The Journal of Black Psychology, 10(2), 63-75.

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

Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. Journal of Consulting Psychology, 23(4), 285-299.

Dean, R.S. (1977). Reliability of the WISC-R with Mexican-American children. Journal of School Psychology, 15(3), 267-268.

Dean, R.S. (1980). Factor structure of the WISC-R with Anglos and Mexican-Americans. Journal of School Psychology, 18(3), 234-239.

Floyd, F.J. & Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. Psychological Assessment, 7(3), 286-299.

Galkowski, T., Pietrulewicz, B., & Scott, R. (1987). Factor analysis of the WISC and WISC-R: A Polish comparison. Journal of Clinical Psychology, 43(5), 504-510.

Geisinger, K.F. (1992). Fairness and selected psychometric issues in the psychological testing of Hispanics. In K.F. Geisinger (Ed.), Psychological Testing of Hispanics. Washington D.C.: American Psychological Association.

Geisinger, K.F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. Psychological Assessment, 6(4), 304-312.

Grubb, H.J. & Dozier, A. (1989). Too busy to learn: A "competing behaviors" explanation of cross-cultural differences in academic ascendancy based on the cultural distance hypothesis. The Journal of Black Psychology, 16(1), 23-45.

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

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

Hakuta, K. & Garcia, E.E. (1989). Bilingualism and education. American Psychologist, 44(2), 374-379.

Heath, S.B. (1989). Oral and literate traditions among black Americans living in poverty. American Psychologist, 44(2), 367-373.

Helms, J.E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? American Psychologist, 47(9), 1083-1101.

Hinkle, J.S. (1994). Practitioners and cross-cultural assessment: A practical guide to information and training. Measurement and Evaluation in Counseling and Development, 27, 103-115.

Irvine, S.H. & Carroll, W.K. (1981). Testing and assessment across cultures: Issues in methodology and theory. In Triandis, H.C. & Berry, J.W. (Eds.).

Handbook of Cross-Cultural Psychology-Volume 2: Methodology (pp.181-242). Boston: Allyn & Bacon.

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

Joreskog, K.G. & Sorbom, D. (1989). Lisrel 7: A guide to the program and applications (2nd ed.). Chicago: SPSS, Inc.

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

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

Kaufman, A.S. (1975). Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16 1/2 years. Journal of Consulting and Clinical Psychology, 43(2), 135-147.

Kaufman, A.S. (1979b). WISC-R research: Implications for interpretation. School Psychology Digest, 8(1), 5-27.

Macmann, G.M. & Barnett, D.W. (1992). Redefining the WISC-R: Implications for professional practice and public policy. The Journal of Special Education, 26(2), 139-161.

Marsh, H.W., Balla, J.R., McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103(3), 391-

410.

Matarazzo, J.D. (1972). Wechsler's measurement and appraisal of adult intelligence: Fifth edition. New York: Oxford University Press.

Mercer, J.R. (1988). Ethnic differences in IQ scores: What do they mean? (A response to Lloyd Dunn). Hispanic Journal of Behavioral Sciences, 10(3), 199-218.

Miller-Jones, D. (1989). Culture and testing. American Psychologist, 44(2), 360-366.

O'Grady, K.E. (1989). Factor structure of the WISC-R. Multivariate Behavioral Research, 24(2), 177-193.

Okazaki, S., & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. Psychological Assessment, 7(3), 367-375.

Pedhazur, E.J. & Schmelkin, L. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Reschly, D.J. (1978). WISC-R factor structures among anglos, blacks, chicanos, and native-american papagos. Journal of Consulting and Clinical Psychology, 46(3), 417-422.

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

Rodriguez, O. (1992). Introduction to technical and societal issues in the psychological testing of Hispanics. In K.F. Geisinger (Ed.), Psychological Testing of Hispanics. Washington D.C.: American Psychological Association.

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

Rousey, A. (1990). Factor structure of the WISC-R Mexicano. Educational and Psychological Measurement, 50, 351-357.

Rychlak, J.F. (1995). Logical learning theory. University of Nebraska Press.

Sandoval, J. (1982). The WISC-R factorial validity for minority groups and Spearman's hypothesis. Journal of School Psychology, 20(3), 198-204.

Sattler, J.M. (1988). Assessment of children: Third edition. San Diego, CA: Author.

Sattler, J.M. (1992). Assessment of children: WISC-III and WPPSI-R supplement. San Diego, CA: Author.

Schiele, J.H. (1991). An epistemological perspective on intelligence assessment among African American children. The Journal of Black Psychology, 17(2), 23-36.

Schmeiser, C.B. (1992). Reactions to technical and societal issues in testing Hispanics. In K.F. Geisinger (Ed.), Psychological Testing of Hispanics. Washington D.C.: American Psychological Association.

Semler, I.J. & Iscoe, I. (1966). Structure of intelligence in negro and white children. Journal of Educational Psychology, 57(6), 326-336.

Silverstein, A.B. (1969). An alternative factor analytic solution for Wechsler's intelligence scales. Educational and Psychological Measurement, 29, 763-767.

Silverstein, A.B. (1973). Factor structure of the Wechsler Intelligence Scale for Children for three ethnic groups. Journal of Educational Psychology, 65(3), 408-410.

Silverstein, A.B. (1977). Alternative factor analytic solutions for the Wechsler Intelligence Scale for Children-Revised. Educational and Psychological Measurement, 37, 121-124.

Slate, J.R. & Jones, C.H. (1995). Preliminary evidence of the validity of the WISC-III for African American students undergoing special education evaluations. Educational and Psychological Measurement, 55(6), 1039-1046.

Stedman, J.M., Lawlis, G.F., Cortner, R.H., & Achterberg, G. (1978). Relationships between WISC-R factors, Wide Range Achievement Test scores, and visual-motor maturation in children referred for psychological evaluation. Journal of Consulting and Clinical Psychology, 46(5), 869-872.

Steele, C.M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist, 52(6), 613-629.

Sternberg, R.J. (1990). Metaphors of mind: Conceptions of the nature of intelligence. New York: Cambridge University Press.

Suzuki, L.A., Vraniak, D.A., & Kugler, J.F. (1996). Intellectual assessment across cultures. In Suzuki, L.A., Meller, P.J., & Ponterotto, J.G. (Eds.) Handbook of multicultural assessment: Clinical, psychological, and educational applications. San Francisco: Jossey-Bass.

Teo, T., & Febbraro, A.R. (1997). Norm, factuality, and power in multiculturalism. American Psychologist, 52(6), 656-661.

Tucker, L.R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38(1), 1-10.

Vance, H.R., Huelsman, C.B., & Wherry, R.J. (1976). The hierarchical factor structure of the Wechsler Intelligence Scale for Children as it relates to disadvantaged white and black children.

Vance, H.R. & Wallbrown, F.H. (1978). The structure of intelligence for black children: A hierarchical approach. The Psychological Record, 28, 31-39.

Wallbrown, F.H., Blaha, J., Wallbrown, J.D., & Engin, A.W. (1975). The hierarchical factor structure of the Wechsler Intelligence Scale for Children- Revised. The Journal of Psychology, 89, 223-235.

Wechsler, D. (1939). Definition and nature of intelligence. In Selected Papers of David Wechsler, 1974, 32-35, New York: Academic Press.

Wechsler, D. (1944). The measurement of adult intelligence. Baltimore: Williams & Wilkins.

Wechsler, D. (1949). WISC Manual. New York: The Psychological Corporation.

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

Wechsler, D. (1955). WAIS Manual. New York: The Psychological Corporation.

Wechsler, D. (1974). WISC-R Manual. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1989). WPPSI-R Manual. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). WISC-III Manual. San Antonio, TX: The Psychological Corporation.

Weiss, L.G., Prifitera, A., & Roid, G. (1993). The WISC-III and the fairness of predicting achievement across ethnic and gender groups. Journal of Psychoeducational Assessment: WISC-III Monograph, 35-42.

Wherry, R.J. (1959). Hierarchical factor solutions without rotation. Psychometrika, 24(1), 45-51.

Appendix A

Within Groups Confirmatory Factor Analyses for Three-Factor Model

Scaled Score Data

	Chi-Square	df	p	Chi-Square/df	TLI	AGFI
White	120.30	24	<.001	5.01	.97	.97
Black	30.47	24	.17	1.27	.99	.96
Hispanic	28.52	24	.24	1.19	.99	.95

Raw Score Data

	Chi-Square	df	p	Chi-Square/df	TLI	AGFI
White	238.31	24	<.001	9.93	.98	.94
Black	31.54	24	.14	1.31	1.00	.96
Hispanic	48.67	24	<.001	2.03	.98	.92

Appendix B

Simultaneous Confirmatory Factor Analyses of Three-Factor Model

Scaled Score Data

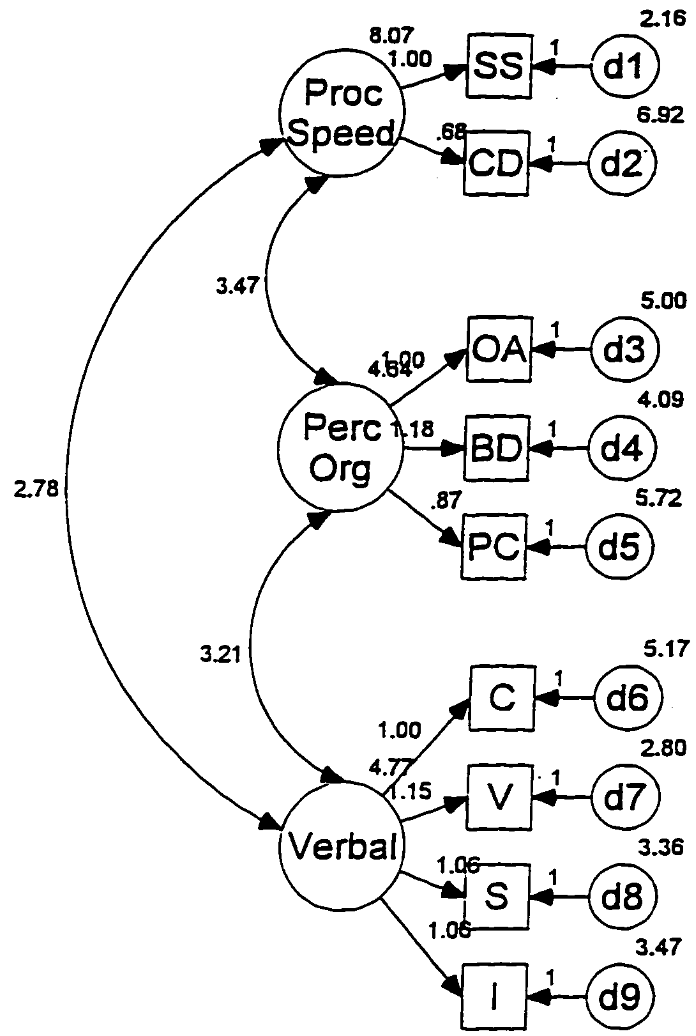
Model	X^2	df	p	X^2/df	TLI	AGFI	ΔX^2	Δdf	p	$\Delta X^2/\Delta$df
3-Factor	179.32	72	<.001	2.49	.98	.96	--	--	--	--
λ	211.63	84	<.001	2.52	.98	.96	32.31	12	<.005	2.69
λ, δ	259.43	102	<.001	2.54	.98	.96	47.80	18	<.005	2.66
λ, δ, Φ	277.91	108	<.001	2.57	.98	.96	18.48	6	<.010	3.08
$\lambda, \delta, \Phi, \Phi_c$	284.15	114	<.001	2.49	.98	.97	6.24	6	NS	1.04

Raw Score Data

Model	X^2	df	p	X^2/df	TLI	AGFI	ΔX^2	Δdf	p	$\Delta X^2/\Delta$df
3-Factor	318.52	72	<.001	4.42	.98	.94	--	--	--	--
λ	360.36	84	<.001	4.30	.98	.94	41.84	12	<.005	3.49
λ, δ	407.34	102	<.001	3.99	.98	.95	46.98	18	<.005	2.61
λ, δ, Φ	431.52	108	<.001	4.00	.98	.95	24.18	6	<.005	4.03
$\lambda, \delta, \Phi, \Phi_c$	435.55	114	<.001	3.82	.99	.95	4.03	6	NS	0.68

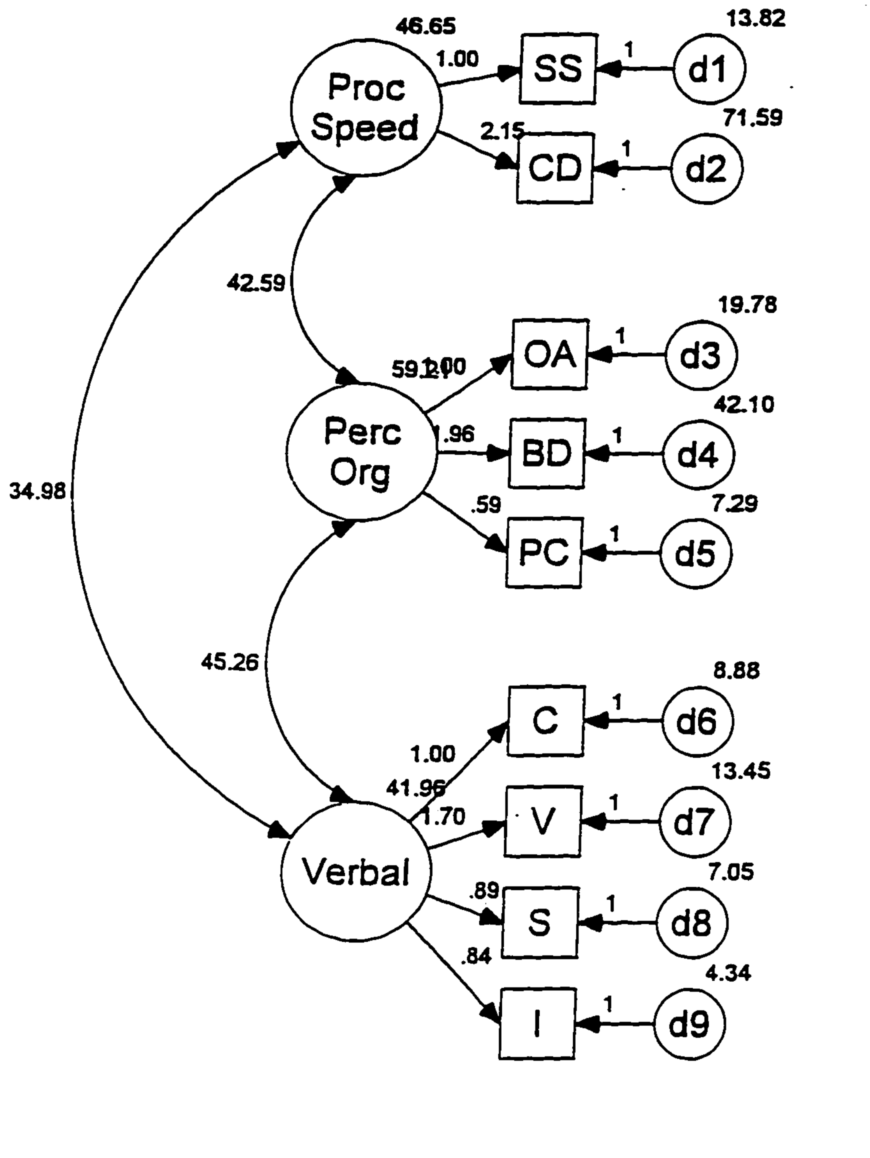
Note. λ = constrained factor loadings; δ = constrained error variances; Φ = constrained factor variances; Φ_c = constrained factor covariances.

Appendix C
Fully Constrained Three-Factor Model: Scaled Score Data



Chi-Square = 284.15 **Chi-Square/df = 2.49**
df = 114 **TLI = .98**
p < .001 **AGFI = .97**

Appendix D
Fully Constrained Three-Factor Model: Raw Score Data



Chi-Square = 435.55 **Chi-Square/df = 3.82**
df = 114 **TLI = .99**
p < .001 **AGFI = .95**

Appendix E

Subtest Unstandardized Factor Loadings for Fully Constrained Three-Factor Model: Raw Data vs. Scaled Score Data*

Factor	Raw Data	Scaled Score Data
Verbal		
Comprehension	1.00 (2)	1.00 (4)
Vocabulary	1.70 (1)	1.15 (1)
Similarities	0.89 (3)	1.06 (2)
Information	0.84 (4)	1.06 (2)
Perceptual Organization		
Object Assembly	1.00 (2)	1.00 (2)
Block Design	1.96 (1)	1.18 (1)
Picture Completion	0.59 (3)	0.87 (3)
Processing Speed		
Symbol Search	1.00 (2)	1.00 (1)
Coding	2.15 (1)	0.68 (2)

*Note. Rank order of the subtest is in parentheses for each factor. The first subtest listed under each factor had its factor loading set to 1.00. This provided a metric on which all other factor loadings for a given factor were based.

Appendix F

Subtest Unstandardized Factor Loadings for White Sample Three-Factor Model: Raw Data vs. Scaled Score Data

Factor	Raw Data	Scaled Score Data
Verbal		
Comprehension	1.00 (2)	1.00 (4)
Vocabulary	1.74 (1)	1.12 (1)
Similarities	0.90 (3)	1.05 (3)
Information	0.86 (4)	1.07 (2)
Perceptual Organization		
Object Assembly	1.00 (2)	1.00 (2)
Block Design	2.00 (1)	1.16 (1)
Picture Completion	0.59 (3)	0.80 (3)
Processing Speed		
Symbol Search	1.00 (2)	1.00 (1)
Coding	2.19 (1)	0.67 (2)

*Note. Rank order of the subtest is in parentheses for each factor. The first subtest listed under each factor had its factor loading set to 1.00. This provided a metric on which all other factor loadings for a given factor were based.

Appendix G

Subtest Unstandardized Factor Loadings for Black Sample Three-Factor Model: Raw Data vs. Scaled Score Data

Factor	Raw Data	Scaled Score Data
Verbal		
Comprehension	1.00 (2)	1.00 (2)
Vocabulary	1.62 (1)	1.14 (1)
Similarities	0.80 (3)	0.90 (4)
Information	0.72 (4)	0.92 (3)
Perceptual Organization		
Object Assembly	1.00 (2)	1.00 (2)
Block Design	1.84 (1)	1.15 (1)
Picture Completion	0.61 (3)	1.00 (2)
Processing Speed		
Symbol Search	1.00 (2)	1.00 (1)
Coding	2.22 (1)	0.78 (2)

*Note. Rank order of the subtest is in parentheses for each factor. The first subtest listed under each factor had its factor loading set to 1.00. This provided a metric on which all other factor loadings for a given factor were based.

Appendix H

Subtest Unstandardized Factor Loadings for Hispanic Sample Three-Factor Model: Raw Data vs. Scaled Score Data

Factor	Raw Data	Scaled Score Data
Verbal		
Comprehension	1.00 (2)	1.00 (4)
Vocabulary	1.57 (1)	1.36 (2)
Similarities	0.90 (3)	1.37 (1)
Information	0.83 (4)	1.16 (3)
Perceptual Organization		
Object Assembly	1.00 (2)	1.00 (3)
Block Design	1.88 (1)	1.39 (1)
Picture Completion	0.59 (3)	1.24 (2)
Processing Speed		
Symbol Search	1.00 (2)	1.00 (1)
Coding	1.87 (1)	0.58 (2)

*Note. Rank order of the subtest is in parentheses for each factor. The first subtest listed under each factor had its factor loading set to 1.00. This provided a metric on which all other factor loadings for a given factor were based.

Appendix I

Subtest Standardized Factor Loadings for Four-Factor Model: Scaled and Raw Score Data

	White		Black		Hispanic	
Factor	Scaled	Raw	Scaled	Raw	Scaled	Raw
Verbal						
Vocabulary	.82	.95	.83	.93	.86	.94
Similarities	.79	.91	.77	.90	.80	.90
Comprehension	.69	.91	.75	.92	.63	.90
Information	.80	.94	.75	.91	.77	.94
Perceptual Organization						
Mazes	.35	.73	.47	.77	.29	.75
Object Assembly	.69	.86	.67	.84	.58	.88
Block Design	.79	.92	.79	.91	.67	.88
Picture Arrangement	.50	.81	.56	.80	.52	.83
Picture Completion	.59	.86	.69	.87	.65	.88
Freedom From Distractibility						
Digit Span	.52	.73	.59	.71	.42	.67
Arithmetic	.83	.95	.71	.90	.72	.92
Processing Speed						
Coding	.61	.88	.64	.84	.45	.80
Symbol Search	.91	.89	.79	.83	.78	.88

Note. All factor loadings are statistically significant ($p < .05$).

Appendix J

Factor Intercorrelations For The Four-Factor Model Within Each Group: Scaled and Raw Score Data

White Group Scaled Score Data

Factor	V	PO	FFD	PS
V	1.0			
PO	.68	1.0		
FFD	.75	.68	1.0	
PS	.43	.59	.51	1.0

White Group Raw Data

Factor	V	PO	FFD	PS
V	1.0			
PO	.92	1.0		
FFD	.92	.90	1.0	
PS	.79	.82	.79	1.0

Black Group Scaled Score Data

Factor	V	PO	FFD	PS
V	1.0			
PO	.74	1.0		
FFD	.86	.75	1.0	
PS	.49	.64	.73	1.0

Black Group Raw Data

Factor	V	PO	FFD	PS
V	1.0			
PO	.91	1.0		
FFD	.96	.93	1.0	
PS	.78	.83	.85	1.0

Hispanic Group Scaled Score Data

Factor	V	PO	FFD	PS
V	1.0			
PO	.78	1.0		
FFD	.91	.80	1.0	
PS	.56	.65	.59	1.0

Hispanic Group Raw Data

Factor	V	PO	FFD	PS
V	1.0			
PO	.93	1.0		
FFD	.98	.96	1.0	
PS	.81	.84	.82	1.0

Note. All correlations are statistically significant ($p < .05$).