

FORDHAM UNIVERSITY
Graduate School of Education

This dissertation prepared under my direction by

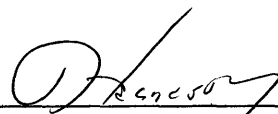
Sharon Shindelman

entitled

Generalizability of the Factor Structure of the WISC-III from
Standardization Samples to African American Students with Learning
Disabilities

has been accepted in partial fulfillment of the requirements for the

Degree of PhD



James J. Hennessy, PhD

4/17/2000

Date



THE
PSYCHOLOGICAL
CORPORATION®

The Psychological Corporation
555 Academic Court
San Antonio, Texas 78204-2498
Tel 210-299-1061
Telex 5106015629 TPCSAT
Fax 210-299-2755

January 20, 2000

Sharon Shindelman
5345 Canvasback Road
Blaine, WA 98230

Dear Ms. Shindelman:

Thank you for your request for permission to extend your agreement for use of the standardization data from the WISC-III in the your study.

This letter will serve as formal authorization to extend the agreement to April 1, 2000 with the same terms and conditions.

Thank you for your continued interest.

Sincerely,

Linda Murphy
Contract Specialist
Legal Affairs

BOOK PERMISSION AGREEMENT

This AGREEMENT entered into as of **SEPTEMBER 15, 1998**, between The Psychological Corporation, a subsidiary of Harcourt Brace & Company, 555 Academic Court, San Antonio, Texas 78204-0952 (herein the "Publisher") and

NAME: **SHARON SHINDELMAN**
ADDRESS: 5345 Canvasback Road
Blaine, Washington 98230

(herein the "Licensee"), WITNESSETH:

WHEREAS the Publisher is the copyright owner of the **WECHSLER INTELLIGENCE SCALE FOR CHILDREN: THIRD EDITION** (herein the "Work"); and

WHEREAS the Licensee wishes to reproduce Tables 1.1, 1.2, and 1.3 from the manual of the Work in Licensee's dissertation concerning urban psychology (herein the "Licensed Uses").

NOW, THEREFORE, the Publisher and the Licensee agree as follows:

1. The Licensee may produce, have produced, and/or distribute such reproductions of the Work specified above, solely for the Licensed Uses and subject to the terms and conditions set forth in this Agreement.

This permission includes non-exclusive world rights in the English language and the right to use any specified illustration(s). Sublicensing may be granted, without charge, for publication or transcription into Braille, large-type editions, or recordings for the blind or other special editions for use by the physically handicapped only if such edition is neither sold nor rented for a fee.

2. The Work shall be identified by title on any reproduction unless otherwise expressly provided in this Agreement.
3. The Licensed Uses specifically exclude the right to print, copy, or distribute in any form, or to translate, adapt, or revise, or to exhibit, represent, record, or reproduce any portion of the Work, either separately or as part of any other larger publication, except as otherwise expressly provided herein.
4. Licensee agrees to delete any portions of the Work which were included in the initial printing of the above-referenced text, but were not permitted herein.
5. All rights in the Work not granted to the Licensee by this Agreement are expressly reserved to the Publisher. The License granted herein relates solely to the first edition of the publication specified in this Agreement.
6. Any reproduction of any portion of the Work shall bear the following notices:

Manual for the Wechsler Intelligence Scale for Children: Third Edition. Copyright © 1991, 1974, 1971 by The Psychological Corporation. Reproduced by permission. All rights reserved.

“Wechsler Intelligence Scale for Children” and “WISC-III” are registered trademarks of The Psychological Corporation.

Notices shall appear on the title page (or reverse side of the title page), of each copy of any reproduction of the Work, or, if the Work is reproduced as part of a larger publication, at the foot of the first page on which the Work is reproduced. Minor rearrangements of the above format may be made in publications for purposes of editorial uniformity, but all the components must be included.

If this Agreement covers more than one Work to be reproduced in one publication, the above model of notice of permission shall be used separately for each separate Work being reproduced, unless a combined form of notice is specifically approved by rider to this Agreement.

7. It is agreed that any FEE SHALL BE WAIVED for the purposes stated herein.
8. The Licensee shall send the Publisher one copy of the Licensed Uses reproduced from the Works under this Agreement.
9. The License granted herein is non-exclusive, and non-transferable to any third party without permission, in writing, from the Publisher.
10. The representative of the Licensee whose signature appears in this Agreement represents and warrants that he has full power to enter into this Agreement.
11. This Agreement shall only become effective if it is executed by the Licensee within thirty days of the effective date shown above.
12. This instrument constitutes the entire agreement between the parties and there are merged herein all prior and collateral understandings and agreements. No amendment or modification of this Agreement shall be valid unless in writing signed by both parties.
13. Regardless of the place of its physical execution or performance this Agreement shall be governed by and interpreted under the laws of the State of Texas, U.S.A.

SHARON SHINDELMAN
Licensee

Sharon A. Shindelman
Name

Graduate Student 10/21/98
Title Date

THE PSYCHOLOGICAL CORPORATION

Aurelio R. Prifitera 10-7-98
Aurelio Prifitera, Ph.D. Date
Vice President

GENERALIZABILITY OF THE FACTOR STRUCTURE OF THE WISC-III FROM
STANDARDIZATION SAMPLES TO AFRICAN AMERICAN STUDENTS
WITH LEARNING DISABILITIES

SHARON ANNE SHINDELMAN, PhD

BA, Seattle University, 1963
MA, Fordham University, 1966

Mentor

James J. Hennessy, PhD

∩

Readers

Vincent C. Alfonso, PhD

Robert M. Thorndike, PhD

DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN
THE GRADUATE SCHOOL OF EDUCATION
OF FORDHAM UNIVERSITY

NEW YORK
2000

ProQuest Number: 13846821

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13846821

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ACKNOWLEDGMENTS

This dissertation was completed as a result of the combined efforts of many individuals and institutions. Dr. Robert M. Thorndike's extraordinary professional competence as a psychometrician, researcher, and teacher, his uncompromising vision of what constitutes research in psychology, and his personal generosity in sharing his time, knowledge, and resources must be acknowledged first. Secondly, I wish to acknowledge the exemplary modeling and teaching of Dr. Sylvia Rosenfield, Dr. Rosa Hagin and Dr. Lisa Suzuki. The support of the Psychological Corporation, especially Dr. Aurelio Prifitera and Cindy Kreiman is gratefully acknowledged. I deeply appreciate the assistance I received from every administrator and support staff member in the school districts that provided data for the AALD sample. Mary Sellazzo, Ryan Murphy, Anita Bechtholt, and Ryann Haw have made invaluable contributions to data evaluation and manuscript preparation. Dr. Meg Cohen, Dr. Carol Franks-Randall and Santina Caruso have provided emotional support over these many years. All of you have shared my dream and helped to make it a reality.

DEDICATION

I wish to dedicate this dissertation to my beloved family, Edward McGraw, Margaret and Cecil Missiaen, and Pamela Duling. The love and support which they have provided me so generously over the many years of my doctoral program, through good times and hard times, has not been acknowledged by me consistently. Without their caring and commitment to this dream and unconditional support this achievement would not have been realized. My love for each of them has deepened through my struggle to reach my goal. They have stayed the course with me and for this I will be eternally grateful to each of them. They always believed in me even when I most doubted myself. I will love you always. You form the most meaningful fabric of my life.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
DEDICATION	iii
LIST OF TABLES	xiii
LIST OF FIGURES	xviii
CHAPTER I. THE PROBLEM	1
The Role of Intelligence Tests	4
Wechsler Intelligence Scale for Children-III	7
Composition of the WISC-III	9
Use of the Index Scores in Test Interpretation	11
Investigations of the Construct Validity of the WISC-III	12
The Contribution of Validity Studies to the use of Cognitive Instruments	12
The Standardization Sample	13
Racial/Ethnic Groups	14
Purpose	15
Definition of Terms	16
Learning Disabilities	17
Race/Ethnicity	17
Research Question	18

TABLE OF CONTENTS (continued)

CHAPTER II. LITERATURE REVIEW	20
The Use of Factor Analysis in Investigating Construct Validity	20
Theoretical Foundations for the Factor Structure of the Wechsler Scales	25
Wechsler's Concept of Intelligence and the Verbal/Performance Dichotomy	25
Kaufman and the Three-Factor Model	29
The WISC-III and the Emergence of the Four-Factor Model	31
Factor Analytic Studies of the WISC, WISC-R, and WISC-III	32
Introduction	32
Factor Analyses of the WISC	32
Factor Analysis of the WISC-R	38
The Factor Analysis of the WISC-R Standardization Sample	38
Investigations of Factor Structure Invariance Across Ethnicity	43
Factor Analysis of the WISC-R Spanish Version	48
Factor Analyses of Clinical Samples Involving the WISC-R	48
Using Factor Analysis to Clarify the Meaning of FD	54
Summary of the Review of Literature on WISC-R Factorization	58
Factor Analyses of the WISC-III	59
Factor Analysis of the Standardization Sample	59
Other Factor Analyses of the WISC-III with Non-Referred Samples	75

TABLE OF CONTENTS (continued)

Factor Analyses of Other National Samples	76
Summary of Studies Involving Children with Learning Disabilities	78
Summary of Studies of Patterns of Abilities	86
A Review of the Literature on Test Bias as it Relates to Validity	90
Content, Standardization Samples and Predictive Validity	91
Examiners' and Language Bias	92
Inequitable Social Consequences	93
Evidence of Construct Bias	97
Qualitatively Distinct Minority and Majority Aptitude and Personality	98
Environmental Issues Involving Assessing the Intelligence of Ethnic Minority Children Living in Poverty	99
Empirical Studies Involving Ethnic Minorities	105
Methodological Issues in Conducting Research on Children with Learning Disabilities	106
Summary	110
CHAPTER III. METHOD	112
Participants	112
The WISC-III Standardization Sample	112
The African American Sample of Children with Learning Disabilities	113
Preliminary Data Treatment of All Samples	116
Instrument	118
Reliability	122

TABLE OF CONTENTS (continued)

Validity	124
Concurrent Validity	124
Construct Validity	124
Procedure	125
Overview of Covariance Structure Modeling	125
Design of the Current Study	133
CHAPTER IV. FINDINGS	141
Introduction and Organization	141
Constraints on Publication of Data	142
Descriptive Statistics	143
Investigation of the Two-Factor Model	147
Overview	147
Partially Restricted Factor Analyses of the Caucasian Sample (CN), the African American Sample (AAN) of the WISC-III Standardization Sample, the Caucasian Validity Sample (CD), and the African American Sample of Children with Learning Disabilities (AALD) for the Two-Factor Model	147
Cross Validation of the Factor Structure of the Caucasian Sample (CN) of the WISC-III Standardization Sample to the African American Sample (AAN), the Clinical Validity Sample (CD) of the WISC-III Standardization Sample, and the Sample of African American Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Two-Factor Model	157
Goodness-of-Fit Statistics	158
LM Test	160

TABLE OF CONTENTS (continued)

Cross-Validation of the Factor Structure of the African American Sample (AAN) of the WISC-III Standardization Sample to the African American Sample of Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Two-Factor Model	162
Goodness-of-Fit Statistics	162
LM Test	163
Cross-Validation of the Factor Structure of the Caucasian Validity Sample (CD) of the WISC-III Standardization Sample to the African American Sample of Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Two-Factor Model	165
Goodness-of-Fit Statistics	165
LM Test	166
Overview of Results for the Two-Factor Model	168
Assessment of Overall Model Fit	168
Investigation of the Three-Factor Model	172
Overview	172
Partially Restricted Factor Analyses of the Caucasian Sample (CN), the African American Sample (AAN) of the WISC-III Standardization Sample, the Caucasian Validity Sample (CD) and the African American Sample of Children with Learning Disabilities (AALD) for the Three-Factor Model	172
Cross Validation of the Factor Structure of the Caucasian Sample (CN) of the WISC-III Standardization Sample to the African American Sample (AAN), the Clinical Validity Sample (CD) of the WISC-III Standardization Sample, and the Sample of African American Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Three-Factor Model	180

TABLE OF CONTENTS (continued)

Cross-Validation of the Factor Structure of the African American Sample (AAN) of the WISC-III Standardization Sample to the African American Sample of Children With Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Three-Factor Model	183
Cross-Validation of the Factor Structure of the Caucasian Validity Sample (CD) of the WISC-III Standardization Sample with Disabilities to the African American Sample of Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Three-Factor Model	185
Overview of Results for the Three-Factor Model	187
Assessment of Overall Model Fit	187
Investigation of the Four-Factor Model	189
Overview	189
Partially Restricted Factor Analysis of the Caucasian Sample (CN), the African American Sample (AAN) of the WISC-III Standardization Sample, the Caucasian Validity Sample (CD) and the African American Sample of Children with Learning Disabilities for the Four-Factor Model	189
Cross Validation of the Factor Structure of the Caucasian Sample (CN) of the WISC-III Standardization Sample to the African American Sample (AAN) of the WISC-III Standardization Sample, the Clinical Validity Sample (CD) of the WISC-III Standardization Sample, and the Sample of African American Children with Learning Disabilities in a Fully Restricted Factor Analysis for the Four-Factor Model	198
Cross-Validation of the Factor Structure of the African American Sample (AAN) of the WISC-III Standardization Sample to the African American Sample of Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Four-Factor Model	200
Goodness-of-Fit	201

TABLE OF CONTENTS (continued)

Cross-Validation of the Factor Structure of the Caucasian Validity Sample (CD) of the WISC-III Standardization Sample to the African American Sample of Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Four-Factor Model	202
Goodness-of-Fit	203
LM Test	204
Overview of Results for the Four-Factor Model	204
Assessment of Overall Model Fit	204
Analysis of the Model Evaluation Statistics Across Samples	206
Comparability of Factor Loadings Across Other Independent Samples	209
Samples of African American Children with Disabilities	209
Samples of Other Ethnically Diverse Samples with Disabilities	209
CHAPTER V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	214
Overview	214
Summary of Statistical Analyses	216
Overview	216
Results for the Two-Factor Model	217
Results for the Three-Factor Model	220
Results for the Four-Factor Model	222
Conclusions	223
The Question of Generalizability of Factor Structure Across Samples	223
The Question of which Independent Variable Causes Greater Model Stress	225

TABLE OF CONTENTS (continued)

The Question of which Model Provides the Best Fit	227
Relating Findings of the Present Study to Other African American Samples with Disabilities	228
Relating Findings of the Present Study to Other Ethnically Diverse Samples with Disabilities	232
Support for the Construct Validity of the WISC-III from the Current Study	235
The Use of Index Scores with African American Children with Learning Disabilities	237
The Use of Index Scores in This Sample of African American Children with Disabilities	239
Substituting Symbol Search for Coding in Calculating Global Scores	240
Limitations of This Study	240
Recommendations for Future Research	242
REFERENCES	244
APPENDICES	
A. The WISC-III	275
Characteristics of the Standardization Sample of the WISC-III	276
Content Revision in the WISC-III	280
B. Studies of Patterns of Abilities of Samples Using the WISC-III Subtest Scores	283
C. EQS Notations and Path Diagrams	295
D. Goodness-of-Fit Indexes Reported in EQS	299

TABLE OF CONTENTS (continued)

E. Selected Fit Statistics for Parallel Partially Restricted Factor Analyses with Constraints for Age Collapsing of African American Children with Learning Disabilities in Two, Three, And Four-Factor Models	305
F. Fit Indices for Fully Restricted Factor Analyses for the Two, Three, and Four-Factor Model	307
G. Standardized Residuals Reported for the Two, Three, And Four-Factor Models for the African American Sample With Learning Disabilities	313
H. Original Correlation Matrix and Standard Deviations for the African American Sample with Learning Disabilities	318
ABSTRACT	320
VITA	323

LIST OF TABLES

Table		Page
1.	The WISC III Subtests Grouped According to Scale	10
2.	Scales Derived from Factor Analyses of the WISC-III Subtests	11
3.	Samples Included in Investigations of the Factor Structure of the WISC-R Across Ethnicity	44
4.	Demographic Characteristics of Clinical Samples Included in Investigations of the Factor Structure of the WISC-R	49
5.	Studies of Patterns of Abilities Generated for WISC-III Scores	89
6.	Description of the WISC-III Subtests	119
7.	Goodness of Fit Indices for Maximum Likelihood Estimation Procedure	131
8.	Methodology: Stages in Cross-Validation Across Ethnicity and Disability Status for All Samples	139
9.	Standard Score Means, Standard Deviations, and Ranges for WISC-III Verbal IQ, Performance IQ, Full Scale IQ, and Subtests for the Entire Sample of African American Children with Learning Disabilities	145
10.	Standard Score Means for WISC-III Verbal IQ, Performance IQ, Full Scale IQ, Verbal Comprehension, Perceptual Organization, Freedom From Distractibility, Processing Speed, and Subtests for Four Clinical Samples of African American Children	146
11.	Structure Coefficients Reported for All Samples in the Partially Restricted Factor Analysis for the Two-Factor Model	149
12.	Model Evaluation Statistics for the Partially Restricted Factor Analysis for the Two-Factor Model	150

LIST OF TABLES (continued)

13.	Factor Loadings for Two-Factor Solutions for All WISC-III Samples and the African American Sample with Learning Disabilities	156
14.	Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Two-Factor Model	159
15.	Model Evaluation Statistics for the Respecified Factor Structure and Loadings of the Caucasian Sample without Disabilities Imposed on the Caucasian Sample with Disabilities and the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Two-Factor Model	161
16.	Model Evaluation Statistics for the Factor Structure and Loadings of the African American Sample without Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Two-Factor Model	163
17.	Model Evaluation Statistics for the Respecified Factor Structure and Loadings of the African American Sample Without Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Two-Factor Model	164
18.	Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample with Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Two-Factor Model	166
19.	Model Evaluation Statistics for the Respecified Factor Structure and Loadings of the Caucasian Sample with Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Two-Factor Model	167
20.	Partially Restricted Factor Analysis for the Three-Factor Model	174
21.	Model Evaluation Statistics for the Partially Restricted Factor Analysis for the Three-Factor Model	175

LIST OF TABLES (continued)

22.	Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Three-Factor Model	181
23.	Model Evaluation Statistics for the Factor Structure and Loadings of the African American Sample without Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Three-Factor Model	184
24.	Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample with Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Three-Factor Model	186
25.	Partially Restricted Factor Analysis for the Four-Factor Model	191
26.	Model Evaluation Statistics for the Partially Restricted Factor Analysis for the Four-Factor Model	192
27.	Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Four-Factor Model	199
28.	Model Evaluation Statistics for the Factor Structure and Loadings of the African American Sample without Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Four-Factor Model	201
29.	Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample with Disabilities Imposed on African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Four-Factor Model	203
30.	Model Evaluation Statistics Across Samples and Models for the Partially Restricted Factor Analyses	207
31.	Factor Loadings for the Subtests Composing the Verbal Comprehension and Perceptual Organization Factors for the African American Sample with Learning Disabilities, Caucasian Validity Sample with Disabilities, and the Kush et al., (in press) Samples in Three-Factor Solutions	210

LIST OF TABLES (continued)

32.	Model Evaluation Statistics for the African American Sample with Learning Disabilities, Caucasian Validity Sample with Disabilities, and the Confirmatory Factor Analysis of the Kush et al., (in press) Samples in the Four-Factor Model	211
33.	Factor Loadings for the Four-Factor Model for the African American Sample with Learning Disabilities, Caucasian Validity Sample with Disabilities, and Konold, Kush, and Canivez (1997) Samples	212
34.	Factor Loadings for Selected Subtests in the Four-Factor Model for the African American Sample with Learning Disabilities, the Caucasian Validity Sample with Disabilities, and the Confirmatory Factor Analysis of the Kush et al., (in press), Samples	213
E1.	Selected Fit Statistics for Partially Restricted Factor Analyses with Constraints for Age Collapsing of African American Children with Learning Disabilities in Two, Three and Four-Factor Models	306
F1.	Model Evaluation Statistics for the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Two-Factor Model	308
F2.	Model Evaluation Statistics for the African American Sample without Disabilities Imposed on the African American Sample with Learning Disabilities for the Two, Three, and Four-Factor Model	309
F3.	Model Evaluation Statistics for the Caucasian Sample with Disabilities Imposed on the African American Sample with Learning Disabilities for the Two, Three, and Four-Factor Model	310
F4.	Model Evaluation Statistics for the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Three-Factor Model	311
F5.	Model Evaluation Statistics for the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Four-Factor Model	312

LIST OF TABLES (continued)

G1.	Largest Standardized Residuals Reported From EQS for the Two-Factor Model for the African American Sample with Learning Disabilities in Factor Loadings for the Caucasian Sample with Disabilities in a Fully Restricted Factor Analysis	314
G2.	Distribution of Standardized Residuals for the African American Sample with Learning Disabilities in the Structures and Loadings of the Caucasian Sample without Disabilities in a Fully Restricted Factor Analysis in the Three-Factor Model	315
G3.	Largest Standardized Residuals Reported From EQS for the Three-Factor Model for the African American Sample with Learning Disabilities in Factor Loadings for the Caucasian Sample with Disabilities in a Fully Restricted Factor Analysis	316
G4.	Largest Standardized Residuals Reported From EQS for the Four-Factor Model for the African American Sample with Learning Disabilities in Factor Loadings for the Caucasian Sample with Disabilities in a Fully Restricted Factor Analysis	317
H1.	Original Correlation Matrix and Standard Deviations for the African American Sample with Learning Disabilities	319

LIST OF FIGURES

Figure		Page
1.	The Sequence of Factor Analyses Across Ethnicity and Disability Status	138
2.	Hypothesized First-Order CFA Model for the Two-Factor Model	148
3.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the Caucasian Sample without Disabilities in the Factor Structure of the Two-Factor Model	151
4.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample without Disabilities in the Factor Structure of the Two-Factor Model	152
5.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the Caucasian Sample with Disabilities in the Factor Structure of the Two-Factor Model	153
6.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample with Learning Disabilities in the Factor Structure of the Two-Factor Model	154
7.	Overall Model Fit of Two-Factor Model Across Samples	170
8.	Respecified Model Fit of the Two-Factor Model Across Samples	171
9.	Hypothesized First-Order CFA Model for the Three-Factor Model	173
10.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the Caucasian Sample without Disabilities in the Factor Structure of the Three-Factor Model	176

LIST OF FIGURES (continued)

11.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample without Disabilities in the Factor Structure of the Three-Factor Model	177
12.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the Caucasian Sample with Disabilities in the Factor Structure of the Three-Factor Model	178
13.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample with Learning Disabilities in the Factor Structure of the Three-Factor Model	179
14.	Overall Model Fit of Three-Factor Model Across Samples	188
15.	Hypothesized First-Order CFA Model for the Four-Factor Model	190
16.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the Caucasian Sample without Disabilities in the Factor Structure of the Four-Factor Model	194
17.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample without Disabilities in the Factor Structure of the Four-Factor Model	195
18.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the Caucasian Sample with Disabilities in the Factor Structure of the Four-Factor Model	196
19.	The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample with Learning Disabilities in the Factor Structure of the Four-Factor Model	197
20.	Overall Model Fit of Four-Factor Model Across Samples	205
21.	Differences in Average Magnitude of the Correlation Matrices of the Sample with Disabilities in Comparison to the Samples without Disabilities	219

CHAPTER I

THE PROBLEM

Children with learning disabilities represent the largest group of school-aged children for whom special education services are mandated by the federal legislation according to the United States Department of Education (Federal Register, March 13, 1999). The Education for all Handicapped Children Act of 1975, commonly referenced as P.L. 94-142, and its successor, The Individuals with Disabilities Education Act, commonly referenced as IDEA, 1999 (Section 300.7 of the Rules and Regulations) defines a learning disability as follows:

“Specific learning disability” means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations, including conditions such as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia ... The term does not include learning problems that are primarily the result of visual, hearing, or motor handicaps, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage. (Federal Register, 1999, p.12422).

The federal definition serves to some extent as the model for each state.

Chalfant (1989) reported all of the state definitions contain two to five of the components embedded in the federal legislation. These components include a task failure component, an exclusion component, an etiological component, a discrepancy component, and a psychological process component (Chalfant, 1989).

Children with learning disabilities are estimated to account for 10-15% of the school age population in the United States (Gaddes & Edgell, 1993; Silver & Hagin, 1990). More than 2.5 million children in the United States are identified as learning disabilities (U.S. Department of Education, 1997 as cited in T. J. Ward, S. B. Ward, Glutting & Hatt, 1999). The rate of classification of children with learning disabilities has experienced its greatest growth in decades subsequent to the passage of P.L. 94-142, with growth rates from 1980-1990 reflecting increases across all racial/ethnic subgroups (McBrayer & Garcia, 1996). McBrayer and Garcia report that during that period the percent change for African Americans was 61% for Caucasians, 59% for Hispanics, 47% for Native Americans, and for Asians and Pacific Americans, 14%. The trend continued into the 1990's. From 1992-1995 alone a 10% increase was reported (U.S. Department of Education, 1997 as cited in T. J. Ward, S. B. Ward, Glutting & Hatt, 1999). In the present century the forecast is that the proportion of minority children "will constitute an ever-increasing percentage of public school students" (Oswald, Coutinho, Best, & Singh, 1999, p. 195).

Although African Americans were reported to represent 13-14% of the population of the United States in the 1988 U.S. census, substantially higher percentages of African American children have been designated as having learning disabilities. Allegations of overrepresentation of minority students, including African Americans, in special education programs for the mildly handicapped has sparked extensive debate for more than three decades at federal and state levels (Morison, White, & Feuer, 1996). Originally the focus of the debate was children classified as educable mentally retarded, but more recently the discussion has broadened to include

children with learning disabilities and emotional disabilities (Bennett, 1983).

Presently, African American children represent 16% of elementary and secondary school enrollments, but they constitute 21% of special education enrollments (Oswald et al., 1999).

Factors related to the patterns of apparently disproportionate representation of minorities in special education have been studied extensively (Harry, 1992, 1994; MacMillan & Reschly, 1998; Oswald et al., 1999; Robertson, Kushner, Starks, and Drescher, 1994). Robertson et al. (1994), examined the ways differences in definitions, methodologies of data collection and analyses, and aggregation of data by state and in national samples impacted upon these factors. They concluded that a major determinant of special education classification, of assignment to disability category, and of placement for culturally and linguistically diverse children is the district in which the child resides. A confluence of factors, which are highly complex and not fully understood, including community demographics, the ethnicity base rate of the school district especially, educational factors including the culture of the school district, and economic variables are relevant (Oswald et al., 1999).

Robertson et al., (1994) reviewed patterns of occurrence of special education enrollments with particular focus on the expanding growth rates of ethnically diverse populations, especially African American school children. Minority children with disabilities are believed to be particularly at-risk for educational failure related to “inappropriate identification, placement, and services” (Oswald et al., 1999, p. 194). A special concern is the concentration of urban school-age children, which is overwhelmingly minority (Oswald et al., 1999). Robertson et al., reported rate of

occurrence by ethnicity and disability category in 15 urban centers (including 15 cities reflecting representation from the Middle Atlantic region, the South, the Southwest, the West and the Midwest, but excluding the Northeast) which substantiated this assertion. In 11 of the 15 cities included in the study, African American students specifically were identified as having learning disabilities at higher percentages than their occurrence in the school districts' populations. In Los Angeles, for example, where African Americans constituted 15% of the student population, 27% of the students identified as having a learning disability were African Americans.

The forecast by Gregg (1995) raises questions about the ability of urban communities to reverse the trend of disproportionate classification of African American children in the near future. The urban ecological context is one wherein increasingly larger numbers of African American children have been living in poverty in the past two decades. Children living in poverty appear to be vulnerable to learning failure (Oswald et al., 1999; Robertson et al., 1994). As noted previously, children living in poverty are also more likely to be placed in special education (McLoyd, 1998; Oswald et al., 1999). McLoyd estimated the likelihood of classification increases 2-3% per year for every year a child lives in poverty.

The Role of Intelligence Tests

The greatest use of intelligence tests occurs in schools, and school psychologists are noted to be their most frequent users (Figueroa, 1990). Braden (1997) reported that school psychologists spend more time administering cognitive assessments than any other kind of assessment. In the assessment process most

commonly utilized in the United States to determine eligibility for classification of a learning disability, a committee on special education must establish that a significant discrepancy exists between what the child should have learned and actually has learned (Jones & James, 1993; Kranzler, 1997; T. J. Ward, S. B. Ward, Glutting & Hatt, 1999). Mercer, C. D., Jordan, Allsopp, and Mercer, A. R., (1996) reported that 98% of the states utilize some form of the discrepancy model in specifying criteria and/or definition for learning disability classification. They reported the four most commonly used procedures for determining a discrepancy include deviation from grade level, expectancy formulae, regression analysis, and standard score comparisons.

Intelligence tests play a prominent role in this process in determining an individual child's relative standing in relation to his/her peers on the dimension of measured intelligence (Matarazzo & Herman, 1985). Intelligence tests are used most frequently as the ability measure or standardized yardstick against which achievement scores are contrasted to determine if an ability/achievement discrepancy exists (Kamphaus, 1993; Kranzler, 1997; Reschly, 1997; Speece, 1994).

Because children with learning disabilities constitute the largest disability category, it appears likely intelligence tests are given more frequently to identify learning disabilities than for any other single purpose (Reschly, 1997). Additionally, intelligence test results often are weighted more prominently than other data in determining learning disabilities (Elliott, S. N. et al., 1985). This is because "although a wide variety of information is usually gathered prior to classification ... the intelligence test results often are regarded by critics as the most important piece of

information” (Reschly, 1981, p. 1095). This importance is evidenced by the predominant role intelligence tests typically assume in written psychological reports (Esters, Ittenbach, & Han, 1997).

The role of intelligence tests is problematic, however, for many reasons. McLeskey and Waldron (1991) reported that criteria for identification were applied inconsistently across local education agencies with respect to test selection, comprehensiveness of assessment, and cutoff scores of both intelligence and achievement measures. There is some evidence that when cutoff scores of 85 or above are used, African American children “are significantly less likely to be identified than white children” (McLeskey, Waldron, & Wornhoff, 1990, p. 364).

Some districts have moved away from strict adherence to discrepancy formulae (MacMillan, Gresham, & Bocian, 1998a). The application of intelligence test scores to the discrepancy model with minorities either when Full Scale Intelligence Quotients or alternative Performance Intelligence Quotients of intelligence tests are used in the discrepancy model reflects “limited congruence” according to MacMillan et al. (1998a). One result of this trend has been the tendency for fewer children to be classified as mentally retarded and more to be classified as learning disabled (MacMillan et al., 1998b).

Prifitera, Weiss, and Saklofske (1998) noted that Hispanic children included in the WISC-III standardization sample often exhibit higher PIQ than VIQ scores. African American children, by contrast, tend to exhibit slightly higher scores on verbal than performance subtests (Prifitera et al., 1998). Prifitera et al. (1998) advocate that agencies using summary measures carefully review in every situation

where classification decisions are being rendered whether these measures represent unitary measures for that particular child.

Calculation of the discrepancy using the IQ score varies greatly across states, however (Frankenberger & Harper, 1987). The most psychometrically defensible method for determining a discrepancy involves regression (Cone & Wilson, 1981; Thorndike, 1963). Although few states have adopted this in their guidelines for determining discrepancies between IQ and achievement (Frankenberger et al., 1987), use of standard score or regression-based procedures in determining eligibility across ethnicity with Caucasian and African American groups produces differential effects by ethnicity (McLeskey et al., 1990; Braden, 1987). African American children were significantly less likely to be identified when cutoff scores of 85 or greater were used and when standard score procedures were employed than when regression-based calculations were used. McLeskey et al., noted that use of regression procedures resulted in a “proportionally balanced representation of black and white students. In contrast, a standard score procedure resulted in the identification of a significantly greater proportion of white students than black student with learning disabilities” (McLeskey et al., 1990, p. 365).

Wechsler Intelligence Scale for Children-III

The most commonly used tool for assessing intelligence and classifying students with learning disabilities has been the Wechsler Intelligence Scale for Children, (i.e., WISC, Chattin & Bracken, 1989; Hutton, Dubes, & Muir, 1992; Reschly, 1997; Silver & Hagin, 1990). The WISC is a downward extension of the Wechsler Bellevue Intelligence Scale published by David Wechsler in 1939 (Sattler,

1992). Since its introduction in 1949 the WISC has undergone two revisions, the Wechsler Intelligence Scale for Children-Revised (WISC-R, 1974), and the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991). Currently the WISC-III is the most widely used psychometric measure of intelligence for children (Donders, 1996; Morrison et al., 1996; Valencia, Rankin & Oakland, 1997). Further, the WISC-III is the most frequently taught psychoeducational assessment instrument, thus assuring it a leading role in the field for the foreseeable future (Alfonso, Oakland, LaRocca, & Spanakos, in press).

The WISC-III has received positive reviews regarding standardization procedures (Sattler, 1992) and overall reliability (Sattler, 1992). Exacting procedures were employed to ensure population-proportionate minority representation at each age level for males and females and balance regarding gender (Weiss, 1993). The removal of item bias in the design of stimulus materials with respect to gender, ethnicity, and geographic regions was another major goal (Wechsler, 1991; Weiss, 1993). Approximately 15.3% of the standardization sample ($n =$ approximately 338) were African American (Wechsler, 1991). About 7% of the standardization sample ($n =$ 154) were classified as having a learning disability, a speech/language disability, an emotional disability, a physical impairment, or eligibility to be enrolled in Chapter 1 compensatory education programs (Wechsler, 1991). Unfortunately, statistical information regarding the test performance of minority children with disabilities, including African American children, was not specified in the WISC-III manual (Kush & Watkins, 1997).

Although African Americans constituted 43.5% of the population in urban centers with populations over 1,000,000, according to the 1988 U.S. Bureau of Census, they were somewhat under-represented in the WISC-III standardization sample. The percentage of the WISC-III standardization sample from metropolitan areas with populations of over 1,000,000 was 36.7 (Wechsler, 1991). Children whose parents had weak literacy skills were difficult to recruit for the WISC-III standardization sample (personal communication with Aurelio Prifitera, Project Director for the WISC-III, April 3, 1995). Written permission for testing was required for the WISC-III in contrast to the WISC-R where verbal permission had been accepted.

Composition of the WISC-III

The WISC-III is comprised of the 12 subtests that were retained from its predecessors, the WISC-R and WISC, and one new subtest, Symbol Search as reported in Table 1. Full Scale, Verbal, and Performance Intelligence Quotients are used most frequently to report test performance. During test construction of the WISC-III, these quotients were calculated utilizing the 11 core subtests, including Symbol Search. Equal weighting was assigned to each core subtest score. Digit Span and Mazes, which are both supplementary subtests, were not included. The supplementary subtests were considered to provide potentially useful information but not contribute as powerfully statistically to intelligence quotients. Mazes, additionally, is rarely given in practice due in part to length of administration. Digit Span is a far weaker predictor of the Verbal Intelligence Quotient than any of the other core subtests.

Prior to the WISC-III revision, factor analytic studies of its predecessors generally supported a two-factor model. Some studies, particularly with the WISC-R, suggested a three-factor model. A fundamental change introduced in the WISC-III involves the use of four factor based index scores (i.e., Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed). The realignment of subtests occurred when the new subtest, Symbol Search (as reported in Table 2), was introduced and unrestricted and restricted factor analyses were conducted utilizing scores from 12 subtests, excluding Mazes (Wechsler, 1991).

Table 1

The WISC-III Subtests Grouped According to Scale

Verbal	Performance
2. Information	1. Picture Completion
4. Similarities	3. Coding
6. Arithmetic	5. Picture Arrangement
8. Vocabulary	7. Block Design
10. Comprehension	9. Object Assembly
^a 12. Digit Span	^b 11. Symbol Search
	^a 13. Mazes

Note. From Manual for the Wechsler Intelligence Scale for Children-Third Edition

(p. 5), by D. Wechsler, 1991, San Antonio: The Psychological Corporation.

Copyright 1991 by The Psychological Corporation. Reprinted with permission.

^aSupplementary subtest. ^bSupplementary subtest that can substitute only for Coding.

Table 2

Scales Derived from Factor Analyses of the WISC-III Subtests

Factor 1 Verbal Comprehension	Factor II Perceptual Organization	Factor III Freedom from Distractibility	Factor IV Processing Speed
Information	Picture Completion	Arithmetic	Coding
Similarities	Picture Arrangement	Digit Span	Symbol Search
Vocabulary	Block Design		
Comprehension	Object Assembly		

Note. From Manual for the Wechsler Intelligence Scale for Children-Third Edition

(p. 7), by D. Wechsler, 1991, San Antonio: The Psychological Corporation.

Copyright 1991 by The Psychological Corporation. Reprinted with permission.

Use of Index Scores in Test Interpretation

The use of the four factor based index scores in test interpretation is recommended by developers of the WISC-III. With respect to evaluation of the performance of certain clinical populations, including children with learning disabilities, specifically, the WISC-III manual (Wechsler, 1991) advocates the use of the Freedom from Distractibility and Processing Speed Indices. The index scores are considered to be more reliable than the scores of the individual subtests from which they have been derived (Donders, 1996; Wechsler, 1991). Prifitera, Weiss, and Saklofske (1998) examined the mean IQs and index scores for the Caucasian, African American, and Hispanic samples within the WISC-III standardization sample. They concluded from their investigation that the finding “strongly supports the practice of

using the index scores even though factor analyses do not always clearly support the four-factor structure for minority groups” (p. 13).

Other researchers, however, raise questions about the interpretability of these index scores based upon several concerns unresolved to date, including lack of psychological theory to support the four-factor conceptualization, lack of research into what effect these factors have on interpretation of assessment protocols, and inconsistent fit of the four-factor model in some age groups in the standardization sample (Carroll, 1993a; Kamphaus, Benson, Hutchinson, & Platt, 1994).

Additionally, researchers suggest there is a lack of findings supporting the four-factor model in clinical populations, including those with learning disabilities (Allen & Thorndike, 1995a; Riccio, Cohen, Hall, & Ross, 1997).

Investigations of the Construct Validity of the WISC-III

The Contribution of Validity Studies to the Use of Cognitive Assessment Instruments

Federal regulations (sections 300.530-300.534 of the abridged rules and regulations for the implementation of P.L. 94-142) require that tests used in evaluating children suspected of having disabilities must demonstrate psychometrically that they have been validated for the purposes for which they are utilized. Test validity, an aspect of construct validity, in its most general applications, is concerned with the evidence that indicates the test measures what it purports to measure. “Validity refers to the number and range of valid inferences a user can make about a client on the basis of test scores” (Most & Zeidner, 1995, P. 493). Construct validity, broadly defined, is the overarching umbrella under which all

aspects of measurement validity are subsumed (Messick, 1989; Most & Zeidner, 1995). The ongoing process of investigating construct validity involves assessing how the construct of intelligence is conceptualized, how it appears in tests and interacts with other behaviors in those situations, and what results are obtained when its hypothesized relations are formally tested (Most & Zeidner, 1995).

The Standardization Sample

Extensive research has been conducted with the WISC-III standardization sample to support construct validity using exploratory and confirmatory factor analytic techniques (Allen & Thorndike, 1995a, 1995b; Carroll, 1993a; Kaufman, 1994; Keith & Witta, 1997; Macmann & Barnett, 1994; Ownby & Carmin, 1994; Sattler, 1992; Wechsler, 1991). Differing conclusions have been reached, however, about the underlying factor structure of the instrument for the entire standardization sample. A one-factor structure is reported by Macmann and Barnett (1994), a two-factor structure by Allen and Thorndike (1995b), a three-factor structure by Sattler (1992), a different three-factor structure by Ownby and Carmin (1994), and a four-factor structure by Keith and Witta (1997) and Wechsler (1991).

The lack of consensus about factor structure of the WISC-III arises from a multiplicity of issues that are beyond the scope of this study. Three are briefly mentioned here: a) the lack of articulation of a strong theoretical foundation prior to test development (Kamphaus et al., 1994); b) the failure of the WISC-III manual to report the actual model, including factor loadings and factor correlations, for the confirmatory analyses (Carroll, 1993b; Keith & Witta, 1997) which has made it impossible for other researchers to replicate these studies precisely and consistently;

and c) the lack of adequate representation of a sufficient number of subtests to define all possible factors (Carroll, 1993b; Gorsuch, 1983; Thorndike, 1990), especially with regards to the Processing Speed and Freedom from Distractibility indices.

Racial/Ethnic Subgroups

Test developers did not report separate analyses within the standardization sample for minority group members. Differing results have been reported about the underlying factor structures in samples of minority students with learning disabilities (Bell, 1994; Konold, Kush & Canivez, 1997; Kush, 1996; Loderquist-Hansen & Barona, 1994; Wechsler, 1991). The Kush (1996) study provided strong support for the traditional Verbal and Performance factors. Loderquist-Hansen and Barona found evidence for a third factor that was identified as measuring processing speed. Bell (1994) and Konold, Kush and Canivez (1997) concluded that the four-factor model that was recommended by Wechsler (1991) had the best fit.

Three studies have been reported to date involving exclusively African American referred samples. The most comprehensive of these studies was conducted by Kush et al. (in press), who investigated the construct validity of the WISC-III for a sample of African American students referred for psychological evaluations. Kush and Watkins (1997) investigated the factor structure of the WISC-III for a sample of African American special education students who were undergoing triennial re-evaluations. Slate and Jones (1995) conducted a study involving exclusively African American students referred for special education.

Three limitations are noteworthy in the three studies reported involving exclusively African American referred samples: a) these studies do not focus

exclusively on the factor structure of African American children with learning disabilities; b) no attempt was made to disentangle the effect of race on factor patterns from the effect of disability category on factor patterns; and c) the three studies used varimax rotations. Varimax rotation does not permit factors to be correlated (Macmann, & Barnett, 1994; Thorndike, 1990), a procedure that is not consistent with the psychological theory of the test as proposed by Wechsler (Macmann, & Barnett, 1994).

The underlying question of whether the structure of cognitive abilities of African American children with learning disabilities differs fundamentally from that of nondisabled children comprising the standardization sample of the WISC-III remains unanswered by investigations conducted to date. In addition to the issue of appropriate application of statistical methodology, Suzuki (1992) identified other methodological problems in the WISC-R research which are relevant to WISC-III investigations.

Results remain inconclusive and inconsistent due to mixed samples of referred and non-referred subjects, small sample sizes, unrepresentative samples and failure to separate out score discrepancies between ethnic groups from ability pattern profiles within ethnic groups. (Suzuki, 1994, p.7)

Purpose

The purpose of this study was to investigate psychometrically whether differences exist in the structure of ability of a sample of African American children with learning disabilities when compared to the Caucasian subset of the WISC-III standardization sample without disabilities. The basic research question under investigation in the present study was: were the factors and loadings of the WISC-III

subtests on these factors stable across these samples? The psychometric approach was chosen to address this question because it has not only inspired the most research and attracted the most attention (to date) but it is by far the most widely used in practical settings (Neisser et al., 1996).

The establishment of psychometric comparability between the test performance of the sample of African American children with learning disabilities in the current study and the Caucasian subset of the standardization sample of the WISC-III without disabilities would provide evidence of the construct validity of the instrument for use with African American children with learning disabilities. It would provide supporting evidence that this clinical sample “perceive(s) and interpret(s) test materials in a similar manner” (Keith & Reynolds, 1990, p. 52) to the standardization sample. It would give test users greater confidence in interpreting results with the special population of African American children with learning disabilities in the manner suggested by the test developers. This is especially important in relation to the WISC-III index scores, which are a new addition to the test, relying upon the verification of the validity of the four-factor model proposed by Wechsler (1991).

Definition of Terms

This study investigated the effects of two variables on the structure of mental abilities of a sample of school age, African American children with learning disabilities, using partially restricted and fully restricted factor analytic techniques. The two main variables in this study were ethnicity and learning disabilities. Methodology employed in conducting this investigation included partially restricted

and fully restricted factor analyses. Definitions relevant to the proposed study are stated as follows:

Learning Disability

For purposes of this study elements of the definition of specific learning disabilities stated in The Individuals with Disabilities Education Act (1991) regulations at Section 300.7 of the Code of Federal Regulations were used. According to these regulations, a specific learning disability is a disorder in one or more of the basic psychological processes which are related to the understanding and utilization of spoken and/or written language which are not the result primarily of handicaps which involve sensory disturbances including visual and auditory difficulties, motor disabilities, mild mental retardation, emotional disturbance, or sociocultural disadvantage. The disorder results in a severe discrepancy between achievement and intellectual ability (Reschly, 1997).

Race/Ethnicity

The category “race/ethnicity” consistent with categories specified within the March, 1988 United States Census survey, included the following groupings: White; Black (African American); native American; Eskimo; Aleut, Asian; Pacific Islander; Other (Wechsler, 1991). For purposes of this study the category Black (African American) was determined by each local school district utilizing data they obtained in their work with the families of the children whose test scores were included in the study.

Research Questions

This study of the WISC-III addressed a basic research question. It investigated the degree to which the factor structure of the test, utilizing whatever model is most commonly employed, generalized across race and ethnicity from the Caucasian subset of the WISC-III standardization sample without disabilities to a sample of African American children with learning disabilities. Three basic models were identified as appearing most frequently in the literature. The first model is a two-factor model with 10 variables. The second model is a three-factor model with 11 variables. The third model is a four-factor model with 12 variables. The two dimensions of differences under study were ethnicity, (i.e., African American or Caucasian), and clinical status, (i.e., without disabilities or with learning disabilities).

For each of the three basic models the following questions were investigated:

(a) was the factor structure stable from the Caucasian WISC-III sample without disabilities to the African American WISC-III validity sample without disabilities? (b) was the factor structure stable from the Caucasian WISC-III sample without disabilities to the Caucasian WISC-III sample with disabilities? (c) was the factor structure stable from the Caucasian sample without disabilities to the African American sample with learning disabilities? Additionally, did the factor structure generalize in the following areas? First, was the factor structure stable from the African American WISC-III sample without disabilities to the African American sample with learning disabilities? Second, was the factor structure stable from the Caucasian WISC-III sample of children with disabilities to the African American sample of children with learning disabilities? Third, when the average change in fit

statistics was compared across ethnicity (i.e., African American WISC-III subset without disabilities and African American sample with learning disabilities) and disability status (i.e., Caucasian WISC-III sample of children with disabilities and African American sample of children with learning disabilities) which independent variable (ethnicity or disability status) was identified as causing more stress in the factor structure. A less important corollary of this research addressed one final question: that is, when the three models were compared across the four samples, which factor model provided the best fit?

CHAPTER II

LITERATURE REVIEW

In order to evaluate the generalizability of the factor structure of the WISC-III to African American children with learning disabilities, it is important to understand several areas of related research. This review of literature addresses the following: (a) the use of factor analysis in investigating the construct validity of tests of intelligence; (b) the history of the evolution of commonly used factor structure models in relation to the Wechsler Scales; (c) a review of factor analytic studies conducted with the WISC-III and its predecessors; (d) a review of the literature on test bias as it relates to validity; (e) issues involving assessing the intelligence of children living in poverty, with particular emphasis on ethnicity; (f) issues in conducting research on ethnic minority children with disabilities; and (g) methodological issues in conducting research on children with learning disabilities.

The Use of Factor Analysis in Investigating Construct Validity

When an intelligence test is employed as an ability measure in evaluating minority children suspected of having learning disabilities, it must be demonstrated that the test is not racially biased. Additionally, it must be shown that the measure has been validated for purposes of decision-making about eligibility for special education services (sections 300.530-300.534 of the abridged rules and regulations for the

implementation of P.L. 94-142). Sabatino, Spangler, and Vance (1995) asserted that accuracy and efficiency in making decisions about eligibility depend partly upon the validity of the instrument utilized to assess intelligence. The examiner makes inferences from the scores a child receives on a general test of intelligence. These inferences raise the following question: Are the conclusions reached based on the scores the child earned appropriate, meaningful and useful?

If the child is a member of an ethnic group different from the majority group upon which the test was standardized, are the inferences made regarding his or her test scores as valid as they would be for a child of the majority group? Keith et al., (1995) assert “the demonstration that a test measures the same constructs across groups is probably the most important critical investigation of bias for a test of intelligence” (p. 347). One aspect of the assessment of construct validity focuses on the known group validity of a test (i.e., how it relates to existing samples that relate to the construct) (Most & Zeidner, 1995). Samples with disabilities have long been identified as groups which require further study and clarification regarding construct validity (Bennett, 1983). Fan, Willson, and Reynolds (1995) assert that statistical investigations of construct validity across ethnic groups can only be conducted reasonably after a substantial research base is available on the assessment instrument.

No single methodology has been identified which accurately assesses construct validity in psychometric instruments (Reynolds & Kaiser, 1990). Factor analysis has been recognized as one of the most important statistical techniques that can be utilized to investigate construct validity (Anastasi, 1976; Cronbach, 1970; Fan et al., 1995; Nunnally, 1978; Thompson & Daniel, 1996). Nunnally (1978) suggests,

“Factor analysis is at the heart of the measurement of psychological constructs” (p. 112-113). The inquiry into the validation of a measuring instrument (e.g., a test of cognitive ability) “demands empirical evidence that the traits purported to be measured are, in fact, the ones actually measured” (Byrne, 1989, p. 504).

Furthermore, in the case of a test of cognitive ability, which would have subscales, “evidence of construct validity is demonstrated if the subscales exhibit a well-defined factor structure that is consistent with underlying theory” (Byrne, 1989, p. 504).

Factor analysis, is a statistical procedure that enables the researcher to identify clusters of items or subtests of a given instrument that correlate highly with one another but less so with other items or subtests, or not at all. Following this, the researcher then clarifies “if patterns of interrelationships of performances among groups of individuals” (Reynolds, 1982, p. 194) emerge.

Although researchers do not necessarily agree that identical factor analyses of an instrument speak to the innateness of the abilities being measured, consistent factor analytic results across populations do provide strong evidence that whatever is being measured by the instrument is being measured in the same manner and is, in fact, the same construct within each group (Reynolds, 1995, p. 559).

Equivalence of measurement across groups is a measurement issue that although rarely tested is crucial in the establishment of the construct validity of an intelligence test (Byrne, 1989). Gorsuch (1983) refers to this statistical phenomenon as “factor invariance.” Allen and Thorndike (1995a), who labeled it “factor structure reliability,” specify that the psychometric utility of intelligence tests “is derived directly from their ability to measure the composition of these factors across age groups and instruments reliably” (p. 648).

The mathematics of factor analysis was developed “to provide mathematical models for the explanation of psychological theories of human ability and behavior” (Harmon, 1967, p. 3). Harmon specifies that the preferred factor solutions in research are those that emphasize both statistical simplicity and psychological meaningfulness. Charles Spearman’s seminal 1904 paper entitled “General Intelligence, Objectively Determined and Measured” and his ensuing research earned him the title of the father of factor analysis (Harmon, 1967). Karl Pearson also is credited with a fundamental contribution in his paper on the method of principal axis in 1901. During the first quarter of this century, Charles Spearman, Cyril Burt, Karl Pearson, Godfrey Thomson, J.C. Maxwell Garnett and Kurt Holzinger were pioneers in developing theories of intelligence and testing them mathematically using factor analytic techniques (Harmon, 1967). Spearman’s theory of intelligence focuses on a strong, unifying, underlying factor of general ability often referred to as “g”. He spent 40 years researching evidence concerning this theory.

During the 1930’s factor analysis was increasingly used to test theoretical assumptions about the nature of intelligence. A paradigm shift occurred, directing focus away from a single underlying general factor, to theories involving group factors. Garnett is credited with generating the concept of multiple-factor analysis. Harmon regards L.L. Thurstone (who may have been the first to label the methodology multiple factor analysis) as having made the greatest contribution in that decade to expanding the mathematical capabilities of the methodology to study the existence of group factors. Thurstone generalized Spearman’s tetrad-difference criterion “to the rank of the correlation matrix as the basis for determining the number

of common factors” (Harmon, 1967, p. 4). This formulation enabled factor analysis to address research questions more comprehensively in evaluating factor matrices, moving away from the focus exclusively on proving the existence of “g” to discovering group factors.

Advances in theory, statistical methods, and the availability of computers to process data have contributed greatly to the ability of researchers to conduct comprehensive factor analytic studies regarding the structure of abilities measured by intelligence tests. Several basic uses are common currently. Factor analytic studies are employed in order to achieve a parsimonious reduction of the variables (subtests) and to reflect the latent traits measured by the tests, utilizing exploratory or unrestricted factor analytic approaches (Hill, Reddon, Jackson, 1985). Factor structures are isolated using exploratory factor analysis “without consideration of the theoretical expectations of the researcher” (Thompson & Daniel, 1996, p. 198). This information is useful in theory building.

Confirmatory (restricted) factor analytic techniques are employed subsequently to determine “if the factor structure of the scale is consistent with the constructs the instrument purports to measure” (Floyd & Widaman, 1995, p. 287). In this situation a researcher “begins with a definite hypothesis (or a number of competing hypotheses) about the factor structure (i.e., the number of factors and the magnitudes of their loadings in each variable)” (Jensen, 1998, p. 69). The confirmatory or restricted factor analysis allows a researcher to test directly the goodness of fit of the hypothetical factor structure of the data.

Thompson (1997) suggests this is useful in three ways. First, rival models can be tested to see if they threaten the credibility of the underlying construct model. Second, this methodology forces precision in defining constructs. Third, it is possible to evaluate different factor analytic models, (i.e., factor structures, for fit as well as for parsimony). Simpler models are preferable and more likely to replicate in later research. Using confirmatory factor analysis, the researcher gathers evidence of the validity of a theory through the investigation of its underlying constructs.

Theoretical Foundations for the Factor Structure of the Wechsler Scales

There is no general consensus that a specific factor model adequately represents the WISC-III. Three models appear in the literature most frequently, a two-factor model, a three-factor model, and a four-factor model. The underpinnings of the Wechsler Scales derive from the intuitive notions of David Wechsler about intelligence and his attempts to construct a cognitive assessment tool that reflected his ideas. His views will be summarized and the origins of those models will be reviewed briefly in the following section.

Wechsler's Concept of Intelligence and the Verbal/Performance Dichotomy

During the early portion of the 20th century, the Stanford-Binet dominated the field of individually administered intelligence testing. It was developed primarily for use with children and adolescents. The Stanford-Binet was derived from the Binet-Simon scale that was developed by Alfred Binet and Theophilus Simon in 1905 (Kamphaus et al., 1997). David Wechsler, a psychologist who assumed a leadership position in the field of clinical psychology during the middle portion of this century,

broadened the focus of intelligence testing to include the assessment of adults with the development of his Wechsler-Bellevue Scale (W-B) in 1939. He constructed this scale to incorporate a more balanced number of verbal and performance subtests because of his concern that the Binet overly emphasized verbal performance in adults (Thorndike, 1997). Items and subtests were borrowed to a large extent from tests constructed earlier in the century (Kaufman, 1979; Thorndike, 1997). The scoring system was borrowed from Yerkes, and the deviation score format from Otis (Thorndike, 1997).

Construction of the test was driven by pragmatic concerns related to his work with clinical populations, particularly at Bellevue Hospital in New York City, including ease and rapidity of administration. Ittenbach, Esters, and Wainer (1997) characterized the Wechsler Scale, like the Binet, as being a “clinician’s test” (p.21), not built upon a theoretical base. Thorndike (1997) described its inception as arising out of a “blunt empirical approach ” (p.14).

Wechsler, in summarizing his views on intelligence, consistently emphasized the global, or overall capacity of the individual to adapt to environmental demands (Lutey & Copeland, 1982; Suzuki, 1992). This emphasis was conceptually close to that of Alfred Binet. Wechsler distinguished his definition and concept of intelligence from that of his colleagues, especially Lewis Terman, however, as containing important differences in two areas:

- (1) It conceives of intelligence as an overall or global entity; that is, a multidetermined and multifaceted entity rather than an independent, uniquely defined trait.
- (2) It avoids singling out any ability, however esteemed (e.g., abstract reasoning), as crucial or overwhelmingly important. In particular, it avoids equating general intelligence with intellectual ability ... Ultimately, intelligence is not a kind of ability at all, certainly not in the same sense that

reasoning, memory, verbal fluency, etc., are manifested under different conditions and circumstances. One can infer an individual's intelligence from how he thinks, talks, moves, almost from any of the many ways he reacts to stimuli of one kind or another (Wechsler, 1974, p.5).

For Wechsler, the concept of intelligence is conceived to be a part of the broader construct of personality. This nonintellective construct includes factors such as persistence and motivation (Wechsler, 1949). He believed these factors, as well as intelligence, could be evaluated with his test.

Wechsler (1974) operationalized this definition of global intelligence in terms of what mental measurement tests represent. Unlike his predecessor, Alfred Binet, Wechsler emphasized shape, not level, in evaluating the meaning of scores on intelligence tests (Kamphaus, Petoskey, & Morgan, 1997):

All tests of ability are essentially set tasks presented to a subject to elicit one or another kind of response that can be readily scored; that is, an artifice so contrived as to permit a subject to communicate meaningfully with an examiner ... To the extent that tests are particular modes of communication, they may be regarded as different languages ... it cannot be assumed that one language is necessarily more valid than another. Intelligence can manifest itself in many forms, and an intelligence scale, to be effective as well as fair, must utilize as many different languages (tests) as possible ... intelligence is best regarded not as a single unique trait but as a composite or global unity (Wechsler, 1974, pp.5-6).

Wechsler (1974) affirmed that the Verbal-Performance dichotomy introduced in the Wechsler-Bellevue generalized across all tests developed later, including the WISC and WISC-R. He explained the dichotomy is "primarily a way of identifying two principal modes by which human abilities express themselves" (p. 9). He added, however, that other ways of classification may be considered. He defended the practice of giving each subtest equal weight in computing Verbal, Performance, and

Full Scale IQ by stating that he subscribed to the theory “that intelligence measures are best regarded as assortative, not hierarchical” (Wechsler, 1974, p. 9).

Curiously, he cautioned that this approach should not be viewed as assuming all subtests contribute equally well to the measurement of intelligence. Rather, it affirms only the necessity of conducting a comprehensive appraisal by including all core subtests (Wechsler, 1974). In fact, Wechsler had explicitly stated in the WISC manual (1949) that, “No attempt has been made to get together a series of tests that measure ‘primary abilities’ or to order them into a hierarchy of relative importance” (p. 5). From this comment one can infer that Wechsler may have been distancing himself from the theory of primary mental abilities proposed by Thurstone in 1938 and from the method of multiple factor analysis which he used to develop this theory.

Factor analytic research related to theories of intelligence and the development of tests to measure intelligence abounded at the time Wechsler developed the Wechsler-Bellevue Scale (W-B) in the late 1930’s. Yet, Wechsler did not utilize this methodology in selecting subtests or organizing them into groups for purposes of defining the Verbal Intelligence Quotient and the Performance Intelligence Quotient in the battery. The Verbal/Performance dichotomy created by Wechsler and used in all subsequent scales he developed was intuitively derived “to represent two different ways that intelligence can be expressed, not two different types of intellectual abilities” (Borgas, 1999, p. 24).

Subsequently, Wechsler did not use factor analysis to investigate the psychometric properties of the W-B or its progeny, the WISC. Although Wechsler mentioned in the 1974 WISC-R manual that factor analytic studies of the WISC had

confirmed the verbal/performance dichotomy of the scale, he included no references to those studies and did not incorporate any factor analyses of the WISC-R in its technical manual. It is ironic that Wechsler left to statisticians the bulk of the research that needed to be conducted to generate evidence of the construct validity of his tests. The factor analytic techniques of these researchers, which Wechsler appeared to ignore, have made it possible for the Wechsler Scales to survive the barrage of criticism from many quarters about test bias and applicability of interpretations from their standardization samples to ethnically diverse clinical populations.

Kaufman and the Three-Factor Model

In 1975 Alan Kaufman, a psychologist with The Psychological Corporation, who worked closely with Wechsler during the development of the WISC-R, published results of his factor analysis of the 11 age groups included in the WISC-R standardization sample. He concluded from these analyses that the WISC-R was best conceptualized as reflecting a three-factor structure (Kaufman, 1975). Consistent with the terms used by Cohen (1959) in his factor analysis of the WISC, Kaufman labeled these factors the Verbal Comprehension factor (VC), the Perceptual Organization factor (PO), and the Freedom from Distractibility factor (FD).

In the model Kaufman promulgated based upon his factor analytic research, VC was defined by four subtests: Information; Similarities; Vocabulary; and Comprehension. PO was defined by five subtests: Picture Completion; Picture Arrangement; Block Design; Object Assembly; and Mazes. FD was defined by three subtests: Arithmetic; Digit Span; and Coding.

Kaufman concluded that “The three factors identified for the WISC-R correspond to meaningful psychological dimensions” (Kaufman, 1975, p. 146). He identified VC and PO as corresponding to Wechsler’s Verbal Scale and Performance Scale, respectively. Kaufman interpreted FD as a measure of what Wechsler had previously characterized as “nonintellective factors” (Wechsler, 1974).

Kaufman suggested a system for measuring the three factors. He recommended the VIQ should be used to represent VC. The PIQ should be used to represent PO. The mean of the subtest scores from Arithmetic, Digit Span, and Coding should serve to represent FD.

In the extensive research using factor analytic techniques to determine the factorial composition of the WISC-R that followed Kaufman’s 1975 study, there was general consensus that Kaufman’s three factor solution had the best fit (Kamphaus, 1993); Reynolds & Kaufman, 1990). Investigation of the third factor revealed it loaded highest on the subtests of Arithmetic and Digit Span and moderately on the Coding subtest (Zachary, 1990). Consensus has not been reached about what this factor actually does measure (Jensen & Reynolds, 1982; Kaufman, 1979; Keith & Witta, 1997; Stewart & Moely, 1983; Wiekiewicz, 1990). From his factor analysis (1993), Carroll concluded the third factor is a combination of a memory span factor and a perceptual speed factor.

Carroll, however, believes the three-factor solution in general is an inadequate interpretation regarding the WISC-R. He based this conclusion on extensive factor analytic studies which he conducted with the Wechsler Adult Intelligence Scale-Revised (WAIS-R), the WISC-R, and the Wechsler Pre-school and Primary

Intelligence Scale-Revised (WPPSI-R). He regarded the WISC-R and WPPSI-R to be essentially downward extensions of the WAIS-R. He concluded all three batteries involved the verbal or language development factor, and the memory span/perceptual speed factor (Carroll, 1993). Additionally, he noted that there is significant subtest specificity in several subtests not accounted for adequately by these three factors. Because these subtests are single measures reflecting possible unique factorial specificity, he concluded “the WISC-R battery is too restricted to permit identification of all factors it measures” (Carroll, 1993, p. 702).

The WISC-III and the Emergence of the Four-Factor Model

One of the goals of the development of the WISC-III was to strengthen the Freedom from Distractibility factor (Wechsler, 1991). In their attempt to accomplish this goal, the test development team incorporated an additional subtest, Symbol Search, into the battery, believing this subtest would enhance FD. To their surprise, when the entire battery was subjected to an exploratory factor analysis, Symbol Search and Coding broke away from FD to form a fourth factor, subsequently labeled Processing Speed (Wechsler, 1991). Arithmetic and Digit Span continued to load on FD (Wechsler, 1991).

Confirmatory factor analyses of subtest scores from the standardization sample were then conducted by the WISC-III test developers. Other investigators (reviewed in greater detail in the subsequent sections on factor analyses of the WISC-III) conducted confirmatory factor analyses as well. These analyses generally supported the four-factor model. As a result of the reconfiguration of the WISC-III subtests into a four-factor structure, according to the test developers, four Index

scores were generated to represent the factor scores. These were incorporated into the scoring system for the WISC-III. Tables for computing Index scores were included in the WISC-III manual (Wechsler, 1991).

Factor Analytic Studies of the WISC, WISC-R, and WISC-III

Introduction

Although this study focused exclusively on factor analytic research involving the Wechsler Scales, it is important to acknowledge other avenues of research that were conducted regarding the WISC. The WISC was the subject of extensive study for nearly three decades. Attempts to evaluate the comparability of the WISC-R revision (same 12 subtests with updated norms and items) to the WISC have been the subject of extensive research over two decades. Shaw, Swerdlik and Laurent (1993) report the “WISC-R was one of the most widely researched assessment instruments” (p. 151). Reynolds and Kaufman (1990) report that most of the 1,100 scholarly articles published since the introduction of the WISC pertain to the WISC-R. Methodological inconsistencies abound across most of the studies of the WISC and WISC-R, making it extremely difficult to evaluate and compare them.

Factor Analysis of the WISC

Cohen (1959) investigated the factor structure of the WISC across three age levels of the WISC standardization sample, 7-6, 10-6, and 13-6. This sample consisted of 200 children (100 males and 100 females) at each age level. The WISC research served as a downward extension of Cohen’s previous factor analyses of the Wechsler-Bellevue (W-B) and Wechsler Adult Intelligence Scale (WAIS). The

matrices of intercorrelations of the three age groups from the WISC standardization sample were analyzed separately using Thurstone's complete centroid method.

Communalities were estimated using Thurstone's Equation 15. Five factors were extracted. The decision to extract 5 factors was determined in large part by the results of the previous factorization of the WAIS. Rotation criteria included "oblique simple structure and a positive manifold with maximization of the number of variables in a plus/minus .05 and plus/minus .10 hyperplane" (Cohen, 1959, p. 286).

After the intercorrelations among the primary factors were determined, they were then subjected to a second-order general factor analysis (Cohen, 1959). From the results of this analysis, the proportion of total and non-error variance attributable to the second-order factor was determined and compared to the proportions found in the WAIS standardization sample. Consistent with the results of the WAIS factorization, five oblique primary factors were found in each WISC age group, when a .20 significance criterion for factor loadings was employed. Overall similarity of factorial composition was found across age groups of children when factor loadings were examined. However, the factor loadings of the subtests were more variable than they had been for the adult groups across the WAIS standardization sample. Error variance was much higher for the WISC age groups than for the WAIS age groups.

The criteria Cohen used to retain factors specified that "at least two exclusively and substantially loading subtests are desirable" (Cohen, 1959, p. 297). He determined three of the WISC factors to be "interpretable" only after he collapsed two splintered verbal factors, which were highly intercorrelated, into a single unified verbal factor. He called the unified factor Verbal Comprehension. Information,

Vocabulary, Similarities and Comprehension were assigned to Verbal Comprehension. The second factor, which he called Perceptual Organization (taken from the WAIS research done previously), reflected consistent and exclusive loadings from Block Design and Object Assembly. Picture Arrangement, Picture Completion, and Mazes loaded less consistently on this factor as well but were assigned to Perceptual Organization based on their factor loadings across some of the age groups.

The third factor, which Cohen labeled Freedom from Distractibility (FD), revealed the consistent and exclusive loading of Digit Span across all 3 age groups. However, Arithmetic loaded on FD for the 13-6 group only. Additionally, Mazes loaded on FD for 10-6 and 13-6, Picture Arrangement on FD for 7-6, and Object Assembly on FD for 10-6. The pattern of loadings of these subtests on FD caused Cohen to conclude that this factor did not rely exclusively on memory, but was also dependent upon other more complex attending cognitive functions, hence the name Freedom from Distractibility.

In Cohen's analyses Coding revealed significant loadings consistently across all 3 age groups on the fifth factor. Picture Arrangement loaded for the 10-6 and 13-6 groups on this factor as well. Cohen labeled this factor "Quasi-Specific" (Cohen, 1959, p. 288). He deemed it to be uninterpretable based upon the fact only one subtest adequately represented it across all age groups. An alternative hypothesis regarding the existence of this factor was proposed by Allen and Thorndike (1995a), who suggested the Quasi-Specific factor may be an artifact of using the centroid method, rendering it not psychologically relevant.

Cohen (1959) found a “substantial degree of correlation among the primary factors” (p. 288) across the three age groups. The second-order general factor emerged (i.e., “g”) when he factored the matrices of factor intercorrelations. This factor accounted for a substantial amount of the variance, but not all of the variance. The correlations of the subtests with “g” are moderate (median average across age groups = .58) suggesting the FSIQ of the WISC is a good measure of “g” (Cohen, 1959). The correlation of the FSIQ with “g” is above .90 across all age groups, as is the VIQ correlation (Cohen, 1959). The PIQ correlations are considerably lower, (i.e., .78, .82, and .81) across the 7-6, 10-6 and 13-6 age groups, respectively. The magnitude of the correlations of the subtests and primary factors with “g” are consistent across all age groups.

Cohen (1959) examined subtest specificity, which is defined as the amount of variance remaining when the proportion of variance shared with other subtests is subtracted from its internal consistency reliability coefficient (Cohen, 1959). He obtained a mean value of .18 and concluded the specificities were too small to serve as a basis for psychometrically defensible pattern analyses and clinical psychodiagnoses. Common variance, by contrast, accounts for slightly over half of the total variance, (.53), “with neither large differences nor a monotonic trend with age” (Cohen, 1959, p. 291). Errors of measurement accounted, on average, for .28 of the total variance and were highest in the youngest age group (Cohen, 1959). Comparable statistical analyses for the WAIS groups yielded higher loadings of general cognitive ability, common variance and lower sampling error because measures of ability are more stable in adults.

Cohen examined the contributions each subtest made to the general factor, subtest specificity, and error variance. Vocabulary and Information were the two strongest measures of “g.” Comprehension, Arithmetic, Similarities, Block Design, and Picture Arrangement were moderately good measures of “g.” The four lowest measures of “g” were Picture Completion, Object Assembly, Coding, and Mazes. Digit Span, because it reflected the largest percentage of measurement error (.40 - .50) was rejected as a measure of “g” (Cohen, 1959).

Lindsey (1967) reported consistency in the factor structure for a sample of African Americans and Caucasians when the two-factor solution was used that reflected the Verbal/Performance dichotomy of the WISC. Silverstein (1973) also obtained consistency in applying the two-factor solution in a sample of Caucasians, African Americans, and Hispanics. Semler and Iscoe (1966) applied a three-factor solution only to a sample of ethnically diverse children and reported significant differences in the second and third factors for Caucasians and African Americans.

Blaha and Wallbrown (1984) used a different theoretical model, adapted from Vernon (1950), to analyze the factor structure of the WISC. Vernon’s model is a hierarchical model, with general intelligence (“g”) at its apex. Two major group factors define the second level, verbal-educational ability (v:ed), and spatial-mechanical-practical ability (k:m). The v:ed factor was defined by positive loadings from all six verbal subtests. The k:m factor was defined by positive loadings from all performance subtests. The second level factors split into minor group factors at a third level. The method of investigation used by Blaha & Wallbrown (1984), is the Wherry and Wherry (1969) hierarchical factor analysis, which involves a principal-

factor solution from the correlation matrix, followed by varimax rotation of the factors. Clusters are then obtained for each factor, using the criterion that each must consist of variables that have the highest absolute loadings on that factor. A “theoretical R matrix (from the relationship $R = FF'$)” (Blaha & Wallbrown, 1984, p. 557) is then constructed. The clusters (with corrected communalities in the diagonals) are next subjected to a Thurstone multiple-group centroid analysis. A cluster intercorrelation matrix is obtained. A factor solution is then obtained on the cluster intercorrelation matrix. This matrix is extended “by appending uniqueness factor loadings for each of the original clusters” (Blaha & Wallbrown, 1984, p. 557). In the final step the extended factor matrix is changed into a transformation matrix, using a modified Newton-Raphson process (Blaha & Wallbrown, 1984). This process alters the transformation matrix so that its transpose becomes its inverse, assuring that the rotations are orthogonal. The altered transformation matrix is then multiplied by the original varimax loadings that were used to develop the theoretical R matrix (Blaha et al.). “This procedure ($W = F \cong TN$) provides the hierarchical factor loadings” (Blaha & Wallbrown, 1984, p. 557).

Blaha and Wallbrown (1984) conducted one study with the WISC standardization sample. Five studies were summarized involving children with reading disabilities, children with mental retardation, and children who were institutionalized. The “g” factor emerged in all but one clinical sample. The two group factors, which corresponded to Wechsler’s V/P dichotomy, emerged with negligible overlap except in two clinical groups, where Kaufman’s FD and VC primary factors replaced the verbal-educative factor. Coding did not consistently help

to define FD, however. A quasi-specific primary factor defined by loadings from Picture Arrangement and Coding appeared in the three samples of children with reading disabilities. Blaha & Wallbrown (1984) concluded Vernon's model did fit the data well across samples overall. Additionally, the FSIQ and V/P dichotomy were reflected in their data.

Factor Analysis of the WISC-R

The Factor Analysis of the WISC-R Standardization Sample

Kaufman (1975) conducted the initial exploratory factor analyses of the scores of the 12 subtests of the WISC-R standardization sample across 11 age levels (n=200 at each age level). Both principal-components factor analysis with orthogonal (Varimax) rotations and principal-axis factoring with squared multiple correlations as initial communality estimates were performed. Subsequently, principal-factor analyses with oblique rotations (oblimax and biquartimin) were performed. The oblique rotations were employed in order to compare results to the previous factorization studies of the WISC by Cohen (1959). Kaufman specified three criteria a factor must meet in order to be meaningful with respect to the WISC-R. First, it must have not fewer than one loading of a minimum .20. Second, it must appear in not fewer than 6 of the 11 age groups. Third, if it did not appear in at least 6 age groups, there needed to be a theoretical rationale why it did not appear. Following these analyses, the patterns of loadings on the unrotated first factor at each age were investigated to clarify the relationship of each subtest to general intelligence. Subtest

specificity for the 12 subtests was determined by subtracting the communality from the reliability, separately by age.

When Kaufman examined the results of the principal-components analysis, he found 2 significant factors at 6 age levels, corresponding to Wechsler's Verbal/Performance dichotomy and 3 significant factors at 5 age levels. At four of these age levels (i.e., 8 1/2, 10 1/2, 13 1/2, and 15 1/2) this factor was similar to Cohen's Freedom from Distractibility factor. At age 14 1/2, however, this factor had highest loadings on Coding and Mazes. Kaufman reported that eigenvalues across the 11 age levels for the third factor "hovered around the 1.0 significance level, ranging from .9 - 1.1" (Kaufman, 1975, p. 137).

The result of the principal-factor analysis, followed by varimax rotations, was examined for two-factor, three-factor, four-factor, and five-factor solutions. Kaufman (1975) determined the results from the three-factor, and to a lesser extent, the four-factor solutions, had the best model fit. The four-factor solution was included because it strengthened the three-factor model in that the three factors emerged consistently across the 11 age levels. In the three-factor model, by contrast, very high loadings by Arithmetic and Digit Span were observed for only 9 of the 11 age levels. At the other two age levels (6 1/2 and 14 1/2), a few performance tests has moderate loadings on the third factor. Coding loaded on it to some extent as well. The pattern of the loadings was similar to the factor Cohen (1959) labeled Quasi-Specific in his investigation of the factorization of the WISC standardization sample.

Using oblique rotations (oblique and biquartimin) Kaufman (1975) reported similar results to those obtained from the orthogonal (varimax) solution. Verbal

Comprehension (VC) and Perceptual Organization (PO) factors again emerged in all 11 age groups. The Freedom from Distractibility (FD) factor emerged again in 9 of the 11 age groups. The quasi-specific factor emerged in the 6 1/2 and 14 1/2 age groups, as it had using the orthogonal solution.

Kaufman (1975) concluded overall that a three-factor structure best characterized the WISC-R standardization sample consistently across both rotational techniques. Kaufman found each subtest to be an excellent measure of a single factor with the exception of Information, which loaded respectably on VC and FD in the varimax solution only.

Kaufman (1975) chose to interpret the large unrotated first principal factor as the measure of general intellectual ability. This measure correlated .90 with the WISC-R FSIQ. The subtests Cohen (1959) found to be the best measures of general intelligence in his factor analysis of the WISC held up as good measures of the WISC-R (i.e., Vocabulary and Information). Additionally, however, Similarities, Comprehension, and Block Design proved to be good measures. Arithmetic, Object Assembly, Picture Completion, and Picture Arrangement were characterized as being fair measures. Picture Completion and Object Assembly were poor measures of general intelligence on the WISC. Consistent with Cohen's findings, Digit Span, Mazes, and Coding proved to be the poorest measures of general intelligence on the WISC-R.

Test specificity measures were stronger regarding the WISC-R subtests than the WISC subtests. The average proportion of specific variance reported for the WISC-R subtests is .28 (Kaufman, 1975); the average error variance is .23 (Kaufman,

1975). Highest specificities across all 11 age ranges were reported for Digit Span, Coding, and Picture Arrangement. Three subtests emerged as having a trivial amount of specific variance, these being Vocabulary, Comprehension, and Object Assembly (Kaufman, 1975). Based upon the fact the average subtest specificity, (i.e., .28) is higher than the average error variance, (i.e., .23) Kaufman defends the clinical practice of profile analysis with the WISC-R. He states “there is enough reliable specific variance to justify some degree of interpretation of an individual’s strengths or weaknesses on the abilities or traits hypothesized for a particular test” (Kaufman, 1975, p. 145).

Although Kaufman (1975) acknowledged that different perspectives existed about the interpretation of FD, he suggested it actually should be placed in what he characterized as the “behavioral domain” (Kaufman, 1975, p. 146). He hypothesized that FD measures what Wechsler (1974) had characterized as nonintellective factors. He offered little evidence to support this hypothesis, however. Even his acceptance of the third factor was questioned by Allen and Thorndike (1995a) because it appeared to violate some of his own basic assumptions about what constitutes a factor in the principal components analysis. Additionally, Allen and Thorndike note the inconsistency of the appearance of the factor in the principal axis factoring.

Blaha and Wallbrown (1984) summarized the results of their hierarchical factor analysis of the WISC-R in 1975, confirming a “g” factor corresponding to FSIQ in the Wechsler model, V/P dichotomy group factors, corresponding to Wechsler’s two-factor model, and minor factors which include Kaufman’s FD, VC, and PO. Their findings were consonant with their investigation of the WISC.

O'Grady (1989) conducted a maximum likelihood confirmatory factor analysis of the 12 subtests of the WISC-R in the 11 age groups of the standardization sample. He fit the data to a one-factor model, both orthogonal and oblique two-factor models, and both orthogonal and oblique three-factor models. In the investigation of the three-factor model, Kaufman's (1979) Freedom from Distractibility model was tested using oblique rotation. Additionally, a variant model was tested with an oblique solution in which Digit Span, Arithmetic, and Coding were permitted to load freely on both the third factor, the first factor (Digit Span and Arithmetic only) or the second factor (Coding only). Subsequently, O'Grady replicated these analyses in 11 clinical samples, which he borrowed from the authors of nine articles published previously.

From his comprehensive reanalysis of standardization sample data, O'Grady concluded the factor structure of the WISC-R is complex and not adequately represented by any of the models specified in the literature. Both models with orthogonal rotations provided considerably poorer fit and were rejected. The oblique multi-factor models provided better fit than the orthogonal models. However, some degree of model misspecification was encountered in both oblique multi-factor models as well as the single-factor model. O'Grady examined the residuals and modification indices for the three models. He concluded no modifications were apparent that would substantially strengthen any of them without adding cumbersome dimensions. He accepted the one-factor solution as providing the best fit because it conformed best to the law of parsimony.

The lack of significant improvement in fit observed in the three factor model in both the standardization sample and the clinical samples caused O'Grady to caution FD may have emerged as a factor in previous studies because of overfactoring. He noted its power of prediction in clinical interpretation would be exceedingly low. Additionally, based upon the results he obtained which caused him to reject the orthogonal models, he concluded that the large body of literature on the WISC-R which was derived from orthogonal rotational methods was based on "improbable underlying models" (O'Grady, 1989, p. 190).

Macmann & Barnett (1994) departed from other interpretations of factor analyses of the Wechsler Scales in reporting results of both exploratory and confirmatory factor analyses of the WISC-R, WAIS-R, and WPPSI-R. Although verbal subtests of the batteries had moderate to high correlations with VIQ and VC (excluding Digit Span, which loaded .59 on the WAIS-R), performance subtests also loaded on VC in the small to moderate range (.36- .63). Macmann & Barnett concluded VC is a "degraded version of the general factor" (Macmann & Barnett, p. 178) and that PO is really an imperfect indicator of the general factor (Macmann & Barnett). They suggest a one-factor solution provides the best fit to the data.

Investigations of Factor Structure Invariance Across Ethnicity

Five investigations of the factor structure of the WISC-R involving samples of children in the United States with regard to ethnicity will be reviewed in this section. A brief summary of demographic characteristics of the samples included in these investigations is included in Table 3.

Vance, Huelsman, & Wherry (1976) examined the factor structure of the WISC for 90 low SES Caucasian and African American children, ages 10-11, whose FSIQ ranged from 80-95, using Vernon's theoretical model. A Wherry hierarchical factor analysis was conducted. The "g" factor did not emerge, according to the authors, because of the restriction set in IQ range. However, the group factors corresponding to Wechsler's V/P dichotomy did emerge across both groups. This sample, labeled by the researchers as "disadvantaged," did perform significantly lower on performance tasks than verbal tasks. Factor structure was consistent across both ethnic groups.

Table 3

Samples Included in Investigations of the Factor Structure of the WISC-R Across Ethnicity

Author(s) & Date	Group ^a	<u>N</u>	Age Range
Vance, Huelsman, & Wherry, 1976.	1, 2	90	10-11
Reschly, 1978.	1, 2, 3, 4	1,040	6-15
Gutkin & Reynolds, 1981.	1, 2	2,175	6-16
Zarske, Moore, & Peterson, 1981.	4, 5	242	6-15
Sandoval, 1982.	1, 2, 6	953	5-15

^a1 = Caucasian; 2 = African American; 3 = Hispanic; 4 = Native American;

5 = Native American Navajo; 6 = Mexican-American

Reschly (1978) evaluated the factor structure of the WISC-R administered to a sample of 1,040 ethnically diverse children from four sociocultural groups, Caucasians, African Americans, Hispanics, and Native American Papagos. Both principal components and unrestricted maximum likelihood analyses were conducted. Two-factor, three-factor, and four-factor models were examined in the maximum likelihood analysis. Subsequently, a principal factor analysis with squared multiple correlations in the diagonals was conducted separately for each group. Varimax rotations of the two-factor, three-factor, and four-factor solutions were then conducted. The two-factor model held up well across all four ethnic groups. The three-factor solution emerged for Caucasians consistently, and Hispanics inconsistently, depending upon which criteria were used. In the Caucasian sample it was identical to the factor structure reported for the standardization sample of the WISC-R by Kaufman in 1975. The three-factor solution did not emerge for African American and Native American children.

Gutkin and Reynolds (1981) examined the factor structure for the Caucasian and African American children included in the WISC-R standardization sample using separate principal factor analyses for two, three, and four-factor models. The two-factor and three-factor models emerged for both groups, although the third factor was described as being weak in the African American sample. For the two-factor model, Coding failed to load substantively on PO. The solution for the three-factor model for both groups resembled Kaufman's three-factor model. Coefficients of congruence that compared unique variances, the general factor, and the two-factor and three-

factor solutions consistently were above .90, indicating factor invariance across ethnic groups.

Zarske, Moore, and Peterson (1981) investigated the factor structure of the WISC-R for a sample of 192 Native American Navajo (129 boys, 63 girls) and 50 Papago Indian children (35 boys, 15 girls). The age range for the sample was 6 years, 8 months to 15 years, 2 months. English was not the primary language of every subject. Principal axis factor analyses were conducted with squared multiple correlations in the diagonals for each group, followed by varimax rotations, specifying two-factor and three-factor solutions. A criterion of .40 was used to identify meaningful factor loadings.

The two-factor solution, for both groups, largely reflected the Wechsler V/P dichotomy. Coefficients of congruence with data from Reschly's (1978) Papago group consistently were .90 or above. Using the criterion of eigenvalues greater than 1.0, Zarske et al. (1981) obtained a different three-factor solution for Navajos than for Papagos. In data analysis of subtest loadings for both groups in the three-factor model, FD did not emerge for either group. The third factor for Papagos was limited to the loading of a single subtest, Digit Span, which was uninterpretable. For the Navajo children, as in Reschly's (1978) African American subset, the third factor involved the splitting of PO into two factors, with Block Design and Object Assembly loading on one factor and Picture Arrangement, Picture Completion, Digit Span, and Object Assembly loading on another.

A principal components analysis was used to determine the extent to which general intelligence estimates were congruent with other samples. Variance estimates

for Navajo children (81%) and Papagos (82%) in this study were similar to Reschly's (1978) findings for African Americans, Papagos, and Caucasians and those of Kaufman for the standardization sample (82%). Zarske et al. (1981), note the results of this factor analysis must be interpreted with caution in light of the small sample size of the Papago group, especially. They concluded, however, the two-factor solution rather than the three-factor solution best characterizes this sample.

Sandoval (1982) examined the factor structure for a sample of 953 children from three ethnic groups who were part of the standardization sample of the System of Multiculture Pluralistic Assessment (Mercer, J., 1979). The sample included 332 Caucasian children, 314 African American children, and 307 Mexican-American children fluent in English. Subtest scores of the WISC-R were factor analyzed for the entire sample and each ethnic subsample. Principal factoring with iterations was performed. Following this, difference scores were calculated for the scores of the Caucasian subsample and the scores of the two minority samples using a formula devised by Sandoval. Subsequently, the comparability of the factor structures generated was investigated. This involved examining the factor vectors. The cosines of the angles formed between factor vectors were calculated and compared. Each measure "was interpretable as a correlation coefficient" (Sandoval, 1982, p.199). The subtest loading for the Verbal and Performance factors were consistent across the three groups, indicating the factor structures were comparable across groups. The three-factor model of Kaufman emerged for the Caucasian subset but not the two minority groups, as had been the case in Reschly's study (1978), reviewed previously.

Factor Analysis of the WISC-R Spanish Version

Gass and Demsky (1998) examined the factor structure of the standardization sample of the Spanish version of the WISC-R, the Escala de Inteligencia para Niños-Revision (EWIN-R; Wechsler, 1982). Data consisted of EWIN-R subtest scores from 532 Cuban American children. Principal components analyses were conducted for 11 age levels incorporating the 11 subtest scores for each child. Overall, data supported the V/P dichotomy of Wechsler, with a few exceptions across the 11 age levels. Kaufman's FD did not emerge consistently in this sample. In this sample Arithmetic and Digit Span had their highest loadings on the Verbal factor in the majority of the age groups. The authors concluded that interpretation of the EWIN-R draws heavily upon the VC and PO factors in use with Cuban American children. FD is not a consistently meaningful factor in this sample.

Factor Analyses of Clinical Samples Involving the WISC-R

Seven studies investigating the factor structure of the WISC-R across clinical samples will be reviewed in this section. Refer to Table 4 for a summary of the demographic characteristics of the samples included in these studies.

Vance and Wallbrown (1978) employed principal-factor analysis using a Wherry and Wherry (1969) hierarchical solution to study the factor structure of the WISC-R for a sample of 150 African American children referred for psychoeducational evaluations. The results of the analysis confirmed the existence of the general factor and the Verbal/Performance dichotomy proposed by Wechsler.

Table 4

Demographic Characteristics of Clinical Samples Included in Investigations of the Factor Structure of the WISC-R

Author(s) & Date	Group ^a	<u>N</u>	Age Range
Vance & Wallbrown, 1978.	1	150	6-15
Petersen & Hart, 1979.	2, 3, 4	594	2 nd through 6 th grade
Karnes & Brown, 1980.	5	946	6-16
Naglieri, 1981.	2	140	6-14
Hale, 1983.	2, 4, 6	265	7-12
Sapp, Chissom & Graham, 1985.	5	371	7-11
Juliano, Haddad, & Carroll, 1988.	2	322	7-16

^a1= African American referred; 2 = Learning Disabled; 3 = Slow Learners;

4 = Emotional Disabilities; 5 = Giftedness; 6 = Mild Mental Retardation.

Petersen and Hart (1979) compared the factor structure of the WISC-R for a clinic-referred population of 594 Utah children to the standardization sample.

All WISC-R subtests except Mazes were administered to all children in this clinical sample. Three subgroups of referred children were factored separately: children with no significant problem (n=248); children with learning disabilities and slow learners (n=162); children with emotional disabilities (n=147). The method of factor analysis

employed, principal components analysis, followed by principal factors analysis, was similar to the one Kaufman (1975) used with the WISC-R standardization sample. Factors 1 and 2 emerged in all three subgroups as the same as Kaufman's Verbal Comprehension factor and Perceptual Organization factor. The loading of Picture Arrangement on Factor 2 for the children with learning disabilities and slow learners was very low (.19). Kaufman's Freedom from Distractibility factor did not emerge for any of these referred groups. The total clinic-referred sample produced a factor structure almost identical to that of the standardization sample "with the exception of a meager loading of Coding on Factor 3" (Petersen & Hart, 1979, p. 645).

Karnes and Brown (1980) examined the factor structure of the WISC-R for 946 children with giftedness, using principal factor analysis. A two-factor solution reflecting Wechsler's V/P dichotomy emerged overall, with two minor exceptions. Arithmetic split its loadings on the two factors. Coding did not load highly on the performance factor. The three-factor solution did not follow Kaufman's three-factor model fully. VC and PO followed the model. FD, however, did not emerge as specified by Kaufman. Instead, Arithmetic and Picture Completion loaded on the third factor for this sample. Coding's loading was much smaller on this factor in this sample (.35). Cosines among factor axes, representing the degree of factor intercovariance, were all above .99. No significant gender differences were found. Karnes et al., (1980) concluded that the factor structure for gifted children is similar to that identified for other special groups.

Naglieri (1981) investigated the factor structure of the WISC-R in a sample of 140 Caucasian children (96 boys, 44 girls), ages 6-2 to 14-8, which had recently been

classified as having a learning disability. First, a principal component factor analysis was conducted on the subtest scores of the 11 subtests administered (excluding Mazes). In the principal component analysis, 1.00's were specified in the diagonal and a varimax rotation was conducted, retaining all factors having eigenvalues greater than 1.00. Following this analysis, a principal axis factor analysis was conducted with squared multiple correlations in the diagonal and the final estimates of the communalities generated through iteration. Wechsler's V/P dichotomy two-factor model, Kaufman's three-factor model, a four-factor model (unspecified), and a five-factor model (unspecified) were examined. Two criteria were used to determine best fit of a model. The first was the number of factors that emerged in the principal component analysis. The second was the empirical rule of accepting only factors with eigenvalues of 1.0 or greater.

Three components were identified in the principal component solution. This solution corresponded partially to Kaufman's three-factor solution. Loadings of subtests on VC and PO were identical. FD was more problematic. Digit Span and Coding loaded moderately, .67 and .75, respectively. Arithmetic only loaded .23 on FD. It loaded slightly higher on VC (.33). Picture Arrangement loaded .29 on FD. It loaded .52 on PO.

In the principal axis factor analysis Wechsler's the two-factor model was confirmed in the two-factor solution (Naglieri, 1981). When the three-factor model was tested a "distractibility factor" emerged. The four-factor solution produced a Verbal, Perceptual and a Distractibility factor as well as an anomalous factor that had only one subtest loading above .30 (i.e., Arithmetic with a loading of .56). The .

uninterpretability of the fourth factor caused Naglieri to select the three-factor model as having the best fit. VC in the three-factor solution replicated Kaufman's VC subtests, as did PO. The composition of FD was more problematic. For this sample Arithmetic did not load on the third factor with Digit Span and Coding. However, Picture Arrangement did load on that factor with a somewhat smaller loading. Based upon this finding, Naglieri speculated that the third factor might measure successive processing or sequencing ability in children with learning disabilities.

Hale (1983) examined the factor structure of the WISC-R for a sample of 265 children with disabilities falling in three categories of disabilities: children with emotional disabilities; children with learning disabilities; children with mild mental retardation. The socioeconomic status (SES) of each child was determined using the criteria of Hollingshead. Two groups were reported: low SES = 158; middle SES = 107. The racial composition of the low SES group included 57% Caucasian and 43% African American children. The racial composition of the middle SES group included 84% Caucasian and 16% African American children. The age range of the sample was 7-12.

The covariance matrices for the low SES group and the middle SES group were generated using scores from the 10 core subtests of the WISC-R. Hale (1983) used the Box procedure to test for equivalence of the covariance matrices. The results indicated the covariance matrices were not significantly different across SES groups. Two interpretable factors emerged in both groups in a principal components analysis conducted separately for each SES group, using Varimax rotation to simple structure. For both groups these factors corresponded to the V/P dichotomy (i.e.,

two-factor model). Coefficients of congruence were computed and were highly congruent across SES. Hale concluded this study suggested the factor structure of the WISC-R was consistent regarding SES across a diverse clinical sample that was racially heterogeneous.

Sapp, Chissom, and Graham (1985) examined the factor structure of the 10 core subtests of the WISC-R for a sample of 371 Caucasian elementary school children (194 boys, 177 girls) who were labeled as “academically superior” (i.e., gifted). Principal axis factor analysis using squared multiple correlations on the diagonal with varimax rotation were conducted, specifying two- and three-factor models. The factor structure of this sample was similar to the Karnes and Brown (1980) gifted sample and Kaufman’s analysis of the WISC-R standardization sample for VC and PO. This sample departed from both samples, however, regarding the third factor. For this sample, Information and Vocabulary had the highest loadings on the third factor. Coding did not load appreciably on any of the factors. Sapp et al., caution that because Digit Span was not administered, the results are not entirely comparable to the other two samples.

Juliano, Haddad, and Carroll (1988) examined the stability of the factor structure of the WISC-R for a sample of 322 children (equally represented by Caucasian and African American subsamples) with learning disabilities over 3 years. Gender representation was equal in both ethnic groups. Scores for the 10 core subtests were available for all 322 children. Scores on Digit Span were available for 229 children. Principal components analyses were conducted for the total sample without Digit Span, for the Caucasian subsample with Digit Span, and for the African

American subsample with Digit Span. Coefficients of congruence were calculated as well.

Results of the factor analysis for the total sample (without Digit Span) were somewhat different from anticipated. A third factor did emerge. Arithmetic split its loadings between Factor 1 and Factor 3 and Coding loaded on Factor 3 rather than Factor 2. In the second investigation for the total sample the two-factor model emerged but Arithmetic split its loadings between VC and PO. When Digit Span was included, Kaufman's three-factor model appeared consistently. Coefficients of congruence across all groups were consistently above .90 for VC and PO. FD coefficients of congruence were above .90 except for the female group (.67). Juliano et al. (1988) concluded the factor structure was invariant across ethnic groups and genders across intervals for this sample of children with learning disabilities.

Using Factor Analysis to Clarify the Meaning of FD

Ownby and Matthews (1985) attempted to clarify what FD measures through a factor analysis of its relationship to selected neuropsychological measures that have been used successfully to assess children with learning disorders. The following instruments were included in the study: WISC-R subtests; Trail Making (Parts A and B); Finger Agnosia Test; Finger Tapping Test of the Halstead-Reitan Battery; Developmental Drawings test of the Wisconsin adaptation of the Halstead-Reitan Battery; Verbal-Spatial test of the Wisconsin adaptation of the Halstead-Reitan Battery; Knox Cube Test; and Peabody Picture Vocabulary Test-Revised (PPVT-R). This battery was administered to 119 children (97 boys and 22 girls, ages 9-14).

These children had been referred for a neuropsychological evaluation because of learning and/or behavior disorders.

Factors derived from the factor analyses using an iterative principal factor solution which were associated with eigenvalues greater than one were then rotated using varimax rotation. The first factor reflected high loadings from neuropsychological tests thought to measure visual-spatial abilities and PO and FD scores from the WISC-R. The second factor had highest loadings from the PPVT-R, VC scores from the WISC-R but also had substantial loadings from PO and FD scores. This factor was interpreted to represent “psychometric intelligence” (Ownby & Matthews, 1985, p. 532). Three other factors emerged with loadings from one measure each. They were characterized as being “method-specific” factors (Ownby & Matthews, 1985, p. 532).

Ownby & Matthews (1985) concluded that measures loading on the first factor reflected more than distraction. They reported these measures tapped visual spatial processing, rapid cognitive processing, and components requiring organizational ability, relating these abilities to executive cognitive processes which children with learning difficulties often display. They concluded their factor analysis of the measures reported above supports the contention that the label FD does not accurately represent the abilities measured by that factor on the WISC-R. The result of this factor analysis must be viewed with caution, however, due to their very small sample size in relation to the large number of measures in the battery administered.

Stone (1992) conducted a joint confirmatory factor analysis of the Differential Ability Scales (DAS; Elliott, C. D., 1990) and the WISC-R. The purposes of this

analysis were threefold: a) to gather evidence of construct overlap and independence; b) to compare competing models; and c) to determine greater generalizability by using a representative rather than a clinical sample. The sample consisted of 115 children (57 boys, 58 girls) who were included in two concurrent validity studies of the DAS. Ethnic composition of the studies included: 92 Caucasians; 9 Hispanics; 9 African Americans; 5 other (Native American, Asian American, and Pacific Islander). Interpretation of findings is problematic because of the small sizes of the ethnic subsamples.

There is some overlap in subtest comparability between the DAS and the WISC-R, especially regarding subtests that load on the Verbal Comprehension factor. There are significant differences, however, in that the DAS has a composite reflecting nonverbal reasoning and a separate composite representing spatial ability. Additionally, subtests involving memory are included as supplementary subtests. They are not used to calculate composite IQ scores at the group factor level or general factor level. This reflects C. D. Elliott's view of intelligence, which is more restrictive than Wechsler's, in that he includes only subtests involving conceptualization and transformation of information in composite IQ scores (Elliott, C. D., 1990).

Five models were compared in Stone's (1992) confirmatory factor analysis. The one-factor model specified all subtests would be forced to load on one factor. The second model, a Wechsler V/P dichotomy model, forced all subtests to load on the Verbal or Performance factors. The three-factor model forced all subtests to load on Kaufman's Verbal, Performance, and Freedom from Distractibility factors. The

latter was defined by “attentional-numerical” content (Stone, 1992, p. 188). The four-factor model conformed to C. D. Elliott’s composite configuration with the supplementary diagnostic subtests forced to load on the Freedom from Distractibility factor. The five-factor model specified C. D. Elliott’s composite configuration, but allowed DAS diagnostic subtests “to form their own separate factors with WISC-R analogues” (Stone, p. 188).

Stone reported three goodness-of-fit indices in addition to the chi-square: the goodness-of-fit index (GFI); the adjusted goodness-of-fit index (AGFI); and the root mean square residual (RMSR). Every model produced better fit than the one-factor model. The five-factor model provided the best fit but even in this model although the RMSR was below .10, an acceptable level, the other two indices were less than .90 (GFI = .85; AGFI = .80). Stone concluded his analyses did not support FD. Stone reported that the analysis suggested that when Coding was allowed to break away from FD and combine with the DAS subtest Speed of Information Processing (similar to the WISC-III Symbol Search) to form a Processing Speed factor, the model fit was improved. Additionally, in this model, Arithmetic and Digit Span broke away from FD and formed a “Numeric Ability” Factor with the DAS subtest Recall of Digits (Stone, 1992, p. 191-192). Because publication of the DAS preceded that of the WISC-III, this is an especially interesting finding. Stone concluded from his analysis that FD should be split, agreeing with evidence from Woodcock (1990). A potentially problematic issue in Stone’s methodology is the reliance upon the GFI and AGFI, both of which have been found to be biased in small samples (Marsh, Balla, & McDonald, 1988).

Using C. D. Elliott's Nonverbal Reasoning and Spatial Ability Clusters, a better fit for the nonverbal subtests was obtained overall. Stone reported Wechsler's Performance subtests loaded only on Spatial Ability, except for Picture Arrangement and Mazes, which had low loadings even on that factor. Both the Verbal and Nonverbal Reasoning factors had the highest loadings on "g" and were highly intercorrelated (.76). Stone concluded that evidence for the Numeric Ability factor, which also had a high "g" loading, argued against Wechsler's and Kaufman's contention that FD (which is composed of two subtests in common with Numeric Ability) is a good measure of the nonintellective factor.

Summary of the Review of Literature on WISC-R Factorization

There is general agreement that the WISC-R reflects factor invariance with respect to Wechsler's V/P dichotomy (two-factor solution) across various ethnic and exceptional groups (Reynolds & Kaufman, 1985). Reynolds and Kaufman, characterize VC and PO as being "robust" across samples of children with learning disabilities. They attribute small differences in factorial composition largely to sampling error or chance fluctuations. The failure of Kaufman's model of FD, which reflects a three-factor solution, to account consistently for subtests that load on the third factor is noteworthy. In several samples, including the EWIN-R standardization sample, the ethnic minority SOMPA subsamples and Reschly's (1978) ethnically diverse subsamples Kaufman's model is not substantiated. Occasionally Picture Completion or Picture Arrangement load on the third factor in exceptional groups. Additionally, Coding does not always emerge as loading on FD. These claims of

evidence lend some support to the suggestion by O'Grady (1989) and Kamphaus et al., (1994) that Kaufman's three-factor model is over factored.

Factor Analyses of the WISC-III

Factor Analysis of the Standardization Sample

The WISC-III manual (Wechsler, 1991) reported a four-factor solution from both exploratory and confirmatory factor analyses of the WISC-III standardization sample. Two major factors and two smaller supplementary factors were reported from analyses of the total sample and four age group subsamples (i.e., ages 6-7, 8-10, 11-13, and 14-16). Inconsistencies were reported only with regard to the 6-7 year old age group with respect to Picture Arrangement, Mazes, and Symbol Search, which "split into two or more factors although they retain loadings of about .30 or more on the targeted factors" (Wechsler, 1991, p. 190).

Sattler (1992), however, reached a different conclusion when he conducted a maximum-likelihood factor analysis (with varimax rotation) for 2, 3, and 4 factor solutions for the entire standardization sample and the 11 separate age groups, in contrast to the four combined age group subsamples used by the Psychological Corporation. He concluded a three-factor solution best characterized the WISC-III, although only a two-factor solution emerged at ages 6 and 15. The three factors he identified were Verbal Comprehension, Perceptual Organization, and Processing Speed. He reported these three factors accounted for 25, 16, and 10 percent of the total variance, respectively, in the three-factor solution (Sattler, 1992).

In the four-factor solution investigated by Sattler (1992) FD did not emerge in 4 of the 11 age groups when the classic Kaiser-Guttman criterion of eigen values of 1.0 or higher was utilized as the minimum level of meaningfulness required. Sattler then conducted a principal components factor analysis and the four-factor solution again did not emerge. Sattler (1992) therefore concluded FD was not substantiated in his analysis. He cautioned about its interpretation in report writing “because of its relative weakness” (Sattler, 1992, p. 1046). Both his analyses and that of Thorndike (1992) provided support for PS, however.

Kamphaus (1993) after reviewing data from the factor analysis of the WISC-III presented in the WISC-III manual (Wechsler, 1991) affirmed the robustness of the Verbal Comprehension factor (VC) and the Perceptual Organization factor (PO) with a caution about Picture Arrangement which loads weakly on PO. He noted that although it was assigned to PO with a factor loading of .37 it also loaded on VC .33 and the Processing Speed factor (PS) .25 (Wechsler, 1991). Lack of evidence of criterion and predictive validity for PS was noted. Kamphaus viewed the composition of the Freedom from Distractibility factor on the WISC-III to be especially problematic in light of the very weak loading of Digit Span (mean loading of .34) on that factor in contrast to the high loading of Arithmetic (mean loading of .73).

Carroll’s review (1993b) of the WISC-III focuses on evidence from factorial analyses he conducted to identify the latent variables underlying the WISC-III subtests. He levels criticism at the test developers of the WISC-III for failing to publish the actual models for the four age groups they used in their confirmatory

factor analyses, including factor loading coefficients and standard errors. This makes it difficult for other researchers to replicate their studies and identify the sources of differences in statistics generated. Carroll's reservations about the construct validity of the WISC-III are, however, primarily an extension of concerns he outlined in his factorial analysis of predecessors of the WISC-III (i.e., the restrictiveness of the battery prevents the statistician from identifying all of the relevant factors being measured by all of the subtests).

Carroll (1993b) cautions the data regarding factor analyses presented in the WISC-III manual does not appear to consider the degree to which the factors are intercorrelated (i.e., to partial out the effects of the intercorrelations and to redistribute the variance among the orthogonally rotated factors). In his statistical analysis, Carroll evaluated the amount of variance each variable shared with other variables in the battery and then determined the degree to which scores on individual subtests of a scale were influenced by unique abilities measured only by that specific scale, as opposed to shared abilities. From this analysis he concluded that the majority of the subtests reflect both communality with their factor and other factors but also moderate specificity (Carroll, 1993b). Carroll suggested subtests reflecting a high degree of specificity would be more likely to depart from profiles predicted by the factor index score in the WISC-III manual.

Carroll (1993b) utilized the same age groupings reported in the WISC-III manual in his factor analyses. The 6-7-age grouping departed from the other three groupings in that its PO factor cluster did not have a high loading on the general factor (-.04) although most subtests with the exception of Coding and Mazes did have

substantial loadings on the general factor. Some loadings on PO, FD and PS for the 6-7-age group produced problematic results, suggesting the index scores may not represent subtest performance accurately both in subtest composition of factor index scores and variability in subtest correlations to the factors. The problematic low loading of Picture Arrangement on PO noted by Kamphaus (1993) for the 6-7 age group, for example, was further analyzed by Carroll who noted most of its variance comes from its loading on the general factor, not its loading on PO. At the 6-7 age level FD is derived primary from Arithmetic and Symbol Search, with much lower positive loadings from Picture Arrangement, Mazes, and Digit Span (Carroll, 1993b). Symbol Search loads only minimally on PS at this age level. Most of the variance contributing to this factor score is derived solely from Coding.

The factor index scores of the three older groups were found to reflect individual subtest performance more accurately than for the 6-7-year-olds. Carroll then assessed the contributions of subtests across all age groups to intelligence quotients. Carroll (1993b) suggests there are limitations in use of factor index scores when the traditional interpretation is applied (i.e., the Verbal Intelligence Quotient (VIQ) is determined largely by VC and the Performance Intelligence Quotient (PIQ) is determined largely by PO). Arithmetic, for example, consistent with results obtained on the Stanford-Binet Intelligence Scale, Fourth Edition (Thorndike, 1990) does not load highly on the VIQ but it does load highly on the general factor, suggesting it does make important contributions to the Full Scale Intelligence Quotient (FSIQ). Thorndike (1990) concluded Arithmetic is factorially complex because it falls between VC and PO, directly upon the general factor. The weak

loading of Coding both on PO (for most age levels) and the general factor, suggesting it contributes little to the FSIQ.

Ownby and Carmin (1994) conducted confirmatory factor analyses of the WISC-III standardization sample, including a replication of the CFA of the WISC-III standardization sample, and tests of models used in previous research with the Wechsler Scales. They examined seven models. The first model was a one-factor model where all 13 subtests load on the factor. The second model was a two-factor model corresponding to Wechsler's V/P dichotomy wherein Symbol Search is loaded on the Performance factor. The third model was a three-factor model corresponding to Kaufman's three-factor model wherein Symbol Search loads on the Performance factor. The fourth model was a revised three-factor model, similar to the previous model, except that Symbol Search loads on the third factor with Arithmetic, Digit Span, and Coding. The fifth model was the four-factor model defined in the WISC-III manual. The sixth model was a revised four-factor model labeled Quasi-Specific (harkening back to Cohen, 1959). In this model VC is intact, PO includes Picture Completion, Block Design, Object Assembly, and Mazes, FD consists of Arithmetic and Digit Span, and the Quasi-Specific factor includes Picture Arrangement, Coding, and Symbol Search. Lastly, the seventh model is a three-factor model of executive function wherein VC and FD are combined into the first factor, PO is the second factor, and the executive function factor is composed of loadings from Arithmetic, Digit Span, Coding, and Symbol Search.

Ownby and Carmin (1994), tested a hypothesis, based upon previous research with the WPPSI-R, WISC-R, and WAIS-R and other tests, that FD and PS actually

measure executive functions. The definition of executive functions used in their research contained two criteria a subtest must meet to be included in the Executive Function factor: 1) it should reflect “considerable cognitive complexity”; 2) it should involve “simultaneous parallel processing” (Ownby & Carmin, 1994, p. 3).

Investigation of the components defining cognitive complexity and involving multiple simultaneous cognitive operations resulted in their specifying the following components: “retrieval from long term memory, holding memory in short term store, and performing logical operations on material in short term store” (Ownby & Carmin, 1994, p.4). From their research, they determined Arithmetic, Picture Arrangement, and Coding met these criteria.

Results of the factor analyses of the models were reviewed and Ownby and Carmin (1994) determined that when the “Executive Function” model was used and Arithmetic and Digit Span were constrained to load on both the Verbal and Executive Function factor that model had the best fit across ages. The notion of cognitive complexity that Ownby and Carmin (1994) attribute to Arithmetic and Digit Span can be considered in light of the results of Keith’s (1997) hierarchical factor analyses of the WISC-III (reported later in this section). Keith reported FD has the highest loading on the general factor (“g”) of the four WISC-III factors with respect to the standardization factor. He does not interpret the memory component, as he labels Digit Span, as being a strong measure of complex mental activity. He fails to differentiate, in his observations at least, that Digit Span is composed of two parts, Digits Forward, and Digits Backward. The latter is often used by neuropsychologists

as a measure of executive function in conducting neuropsychological assessments. It is considered to involve complex operations and simultaneous processing operations.

Ownby and Carmin (1994) caution that the interpretability of PS may vary by age and “may be only present in older children” (p. 2). They note different subtests do load on that factor at different ages. This conclusion is supported by Carroll (1993b) as well in his analysis of data from the WISC-III manual (Wechsler, 1991).

Consistent with the findings of Blaha and Wallbrown (1996), Carroll (1993b), and Sattler (1992) their findings suggest FD as defined in the WISC-III manual (Wechsler, 1991) may be valid only for older children. Both assertions make the use of index score interpretation problematic in relation to children’s test performance on the WISC-III.

Three factor models, Wechsler’s V/P dichotomy two-factor model, Kaufman’s three-factor model, and the four-factor model reported in the WISC-III manual (Wechsler, 1991), were subjected to maximum-likelihood confirmatory factor analysis for the WISC-III standardization sample by Kamphaus, Benson, Hutchinson, and Platt (1994). They used the data from the correlation matrices of the 11 age groups reported in the WISC-III manual (Wechsler, 1991). Each subtest was forced to load on only one factor. Factor correlations and factor variances were freely estimated. Initially the likelihood ratio chi-square statistic, degrees of freedom (df) for the model and associated probability value were examined. Other fit indices were then reviewed including the ratio of the chi-square to its degrees of freedom (df), the root mean square residual (RMSR), the goodness-of-fit index (GFI), the Tucker-Lewis index (TLI), and the cross-validation index (CVI).

The results of the analyses suggested the three and four-factor models fit about equally well for the 6 and 9-year age groups. None of the models worked well consistently across the other age groups. The TLI and CVI do tend to favor the four-factor model, however. Mazes, Digit Span, and Coding consistently provided the lowest factor loadings across the three models. Correlations among latent factors were high across all models, ranging from .66 - .86, causing Kamphaus et al., to criticize the WISC-III test development team for reporting orthogonal rather than oblique rotations in the WISC-III manual (Wechsler, 1991). Kamphaus et al., (1994) caution that the four index scores cannot be applied across every age group with the same degree of confidence because of variations in the way the four-factor model fits across the 11 age groups.

The reliability of the factor structure of Wechsler's V/P dichotomy model was examined across the 3-16 age range of the WPPSI-R and WISC-III standardization samples by Allen and Thorndike (1995a). The reliability of the factor structure in the study was limited to a test of the two-factor solution. The decision to use this model was based on two considerations. First, there was general consensus these were the two factors the tests shared. Second, this is the only model across the studies of factorization of the Wechsler Scales that generally has been endorsed by researchers. Allen & Thorndike (1995a) used a restricted version of a methodology called cross-validation of covariance structure models, which they developed, employing unrestricted and restricted factor analyses. This methodology was developed to avoid the subjectivity inherently evident in "eyeballing" factor structures to assess for invariance.

Initially an unrestricted maximum-likelihood factor analysis was conducted with oblique rotation for each specified age group for the WPPSI-R and WISC-III. The pattern coefficients for the common subtests on the two factors and the correlation between both factors were then entered as fixed parameters in a fully restricted factor analysis. The other parameters, including factor variances and error variances were estimated by the structural equation modeling program, EQS-EM386 (Bentler, 1989) Version 4.01 for Microsoft Windows.

The model was then applied back onto the same age group correlation matrix and that of other closely allied age groups across both instruments, in order to provide a comparative basis for the goodness-of-fit indices used. The age pattern for younger children was tested on a minimum of three older age groups. For older children even more comparisons were made. This was systematically done to look for evidence of possible age-related changes as well as factor structure reliability.

Best fit for the goodness-of-fit indices was provided consistently when models were applied back upon the age group from which the data was generated initially. There was, however, “substantial consistency (save for random sampling variation) across all age groups” (Allen & Thorndike, 1995a, p. 16). Three goodness-of-fit indices, the Bentler-Bonnet Normed Fit index, McDonald Fit index, and the LISREL AGFI Fit index consistently were between .85 - .90. The others, including the Bentler-Bonnet Nonnormed Fit index, the comparative fit index, The Bollen Fit index, and the LISREL GFI index were above .90. The Root Mean Squared Residuals, by contrast, which yield low values when models provide acceptable fit, consistently were below .10. Allen & Thorndike (1995a) concluded Wechsler’s V/P

dichotomy two-factor model was invariant within and across the two instruments.

Because there is a tendency for the model fit to degrade as the two-factor model is applied to older age groups the authors suggest “there may be a gradual developmental progression from a two-factor to a three-factor structure in the WISC-III when Symbol Search is not included” (Allen & Thorndike, 1995a, p.17). This hypothesis was tested in the research described in the following review.

Allen and Thorndike (1995b) investigated the stability of the three-factor structure of the latent variables conforming to the Verbal Comprehension factor, Perceptual Organization factor, and Freedom from Distractibility factor across the WISC-III and WAIS-R standardization samples. Selected age groups were drawn from the WISC-III standardization sample including ages 6, 8, 10, 12, 14, and 16. One hundred males and 100 females at each age level were investigated in this study. Nine age groups were represented in the WAIS-R standardization sample, 16-17 ($n = 200$), 18-19 ($n = 200$), 20-24 ($n = 200$), 25-34 ($n = 300$), 35-44 ($n = 250$), 45-54 ($n = 250$), 55-64 ($n = 160$), 65-69 ($n = 160$), and 70-74 ($n = 160$).

The correlation matrices of 11 similar subtests in both instruments were used. Symbol Search from the WISC-III was excluded because there were no comparable subtests on the WAIS-R that could be investigated. From a review of the literature, Allen and Thorndike (1995b), determined the composition of VC for both instruments would involve Information, Similarities, Comprehension, and Vocabulary subtests from both instruments. Block Design, Object Assembly, Picture Completion, and Picture Arrangement were included to determine PO on the WAIS-R. Those four subtests and Mazes defined PO on the WISC-III. For the WAIS-R, FD included

Arithmetic, Digit Span, and Digit Symbol. Arithmetic, Digit Span, and Coding were the three defining subtests for the FD on the WISC-III.

Initially, an unrestricted maximum-likelihood factor analysis with an oblique rotation was conducted for each age group specified above. The one exception was the 25-34 age group for the WAIS-R. Principal axis factor analysis was used for that group when maximum-likelihood failed to converge. Following this unrestricted analysis, Allen and Thorndike (1995b) conducted restricted factor analyses using maximum-likelihood parameter estimates. The same structural equation modeling program was employed as that used in their WPPSI-R/WISC-III factorization study. The same goodness-of-fit indices were computed from this analysis as well.

As in the previous study, the unrestricted factor analytic solution for each age group was used as the model in the series of restricted analyses that followed. The pattern coefficients for each subtest on each factor were entered as fixed parameters to define the model. The intercorrelations among the three latent variables also were specified. The program freely estimated the other parameters, including factor and error variances. Subsequently, for each age group, the identical pattern coefficients and interfactor correlations were imposed back upon that group, to compare goodness-of-fit indices. The model was then imposed on several older age groups as a “test of factor invariance both within and across instruments, and because the models were tested systematically from younger to older age groups, it also provided the opportunity to examine any systematic intellectual changes that might be present” (Allen & Thorndike, 1995b, pp. 652-653).

When WISC-III age group models were imposed upon WAIS-R groups, certain adjustments were made to address issues of the non-equivalent subtests (i.e., WAIS-R Digit Symbol, and WISC-III Coding and Mazes). With respect to Digit Symbol, the three latent variables were left free to be estimated. Coefficients for Coding and Mazes were dropped from the WISC-III models when other coefficients were imposed. On WAIS-R age groups, mean goodness-of-fit indices for the fit indices generally reflected values above .90. The root-mean-square values were the one exception, because best fit here involved smaller numbers. Values were below .10 generally for that index. The best values were obtained when the fully restricted models were imposed back upon the age group data from which they were generated in the unrestricted analysis. However, across the study, the results show “remarkable consistency in the models across age groups and instruments” (Allen & Thorndike, 1995b, p. 656).

These investigators account for the inconsistent appearance of the three-factor model, particularly with respect to subtest structure coefficients, as largely due to sampling fluctuations. “More specifically, it seems a strong possibility that the seemingly discrepant composition of the Freedom from Distractibility factor, particularly in unrestricted factor patterns, may be due to a hypersensitivity of common rotational procedures to sampling fluctuations” (Allen & Thorndike, 1995b, p. 656).

Blaha and Wallbrown (1996) conducted a hierarchical factor analysis of the WISC-III standardization sample using the intercorrelation matrices for the four age groups combined for analysis in the WISC-III manual (Wechsler, 1991) (i.e., 6.5-7.5;

8.5-10.5; 11.5-13.5; 14.5-16.5). A two-factor solution was specified in order to compare results to the WISC-R data gathered by Wallbrown, Blaha, and Engin (1975). A four-factor solution was specified to compare results to the four-factor index scores reported in the WISC-III manual (Wechsler, 1991). Additionally, a higher order general factor was extracted for all factor solutions.

The magnitude of the loadings of WISC-III subtests on the general factor for the two-factor solution and four-factor solution “remained relatively stable across the four combined age groups” (Blaha & Wallbrown, 1996, p. 215). The minor variations that occurred were attributed to sampling errors. Largest loadings on the general factor for VC were consistent with finding from prior WISC-R and WISC analyses, but different from the report in the WISC-III manual (Wechsler, 1991) that FD had the highest loading. When the contributions of the general factor to all variance for two-factor solutions were compared to the WISC-III (“g” = .34), WISC-R (“g” = .36), and the WISC (“g” = .33), remarkable congruence was noted, supporting the construct validity of the WISC-III as a measure of general intelligence (Blaha et al., 1996).

Congruence with Vernon’s (1950) hierarchical model involving an overriding general factor and two subordinate group factors (i.e., verbal-numerical-educational [v:ed] and spatial-mechanical-practical [k:m]) was reported by Blaha et al., in the two-factor solution. The breakout of subtest loadings on the two factors was congruent with the V/P dichotomy with some exceptions. Digit Span, for example, did not contribute significantly to the v:ed factor and Picture Completion tended “to split its loadings between the v:ed and k:m factors at age 8 and above” (Blaha et al.,

1996, p. 215). Additionally, Arithmetic's loading on verbal was lower than the other verbal subtests (although still significant). These findings were consonant with Keith's (1997) and Carroll's (1993b) analyses.

In the four-factor solution, the four subtests defining VC reported in the WISC-III manual (Wechsler, 1991) loaded moderately across all four age groups and were identical to the variables defining VC in the WISC and WISC-R for Blaha and Wallbrown (1984). Block Design and Object Assembly largely defined PO in this study and the composition of the factor was congruent with earlier WISC-R and WISC studies. FD, which had been defined by Arithmetic and Digit Span in previous studies regarding the WISC-R and WISC, did not evidence itself in the 6-7 age group, consistent with Sattler's (1992) findings. Blaha & Wallbrown (1996) note that earlier factor analyses of the WPPSI did not find evidence of FD at younger age levels, suggesting a developmental age consideration in interpreting this factor, echoed in the research of Allen & Thorndike (1995a, 1995b), Carroll (1993b), Ownby and Carmin (1994), and Sattler (1992). Blaha & Wallbrown (1996) report Coding and Symbol Search have moderate to high factor loadings across the four age groups with the exception of 6-7, where Coding loaded only .26.

Blaha & Wallbrown (1996) indicate that subtest specificity for the 13 subtests varied considerably, from a low of .10 for Arithmetic to a high of .56 for Digit Span. The average subtest specificity was .28. Other higher subtest specificities included Mazes (.53), Picture Completion (.32), and Comprehension (.26). Blaha & Wallbrown (1996) suggested subtests reflecting highest specificities can be interpreted more confidently as single subtests when they deviate from scores on

other subtests because of their high specificity. This observation is consonant with Carroll's (1993b) comment about the likelihood they could depart from index score profiles.

Keith and Witta (1997) conducted hierarchical and cross age confirmatory factor analyses of data from the correlation matrices and standard deviations of the standardization sample of the WISC-III. Their research questions involved determining whether the battery measures the same constructs across its entire age span and to clarify the nature of what those constructs are. Comparison of the correlation matrices for the 11 age groups utilizing LISREL multi-sample analysis with several fit statistics (the Differential Fit Value, the Tucker-Lewis index, Bentler's comparative fit index, and the Parsimonious Fit index) suggested that the model fit the data well. The covariance matrices were very similar. Keith & Witta concluded the same constructs were being measured across the 11 age levels.

Step-wise analyses of what the WISC-III measures were introduced, specifying the most stringent replication, initially, with a freeing of some parameters in subsequent analyses. In the most restrictive initial analysis it was specified that first and second-order factor loadings and unique and error variances would be identical across the 11 age groups. This analysis produced a nonsignificant chi-square, suggesting the hypothesis could not be rejected. In subsequent analyses, parameters were allowed to vary but these calculations did not improve the fit of the models. From these analyses Keith & Witta concluded the four-factor solution did provide an excellent fit with extremely consistent results across the 11 age levels.

A subsequent hierarchical confirmatory factor analysis was conducted across 11 age levels in which all first-order factor loadings were significant and most were reported to be “quite high” (Keith & Witta, 1997, p. 97). Across the four factors, four subtests loaded highly: Vocabulary on VC (.85); Block Design on PO (.81); Arithmetic on FD (.82); and Symbol Search on PS (.90). Loadings on VC by its four subtests consistently were in the .70 -.80 range. Loadings on PO were the most variable with Picture Arrangement (.55) and Mazes (.38) loading the lowest on PO. The loading of Digit Span on FD (.52) was significantly lower than the loading for Arithmetic. The loading of Coding on PO (.59) was significantly lower than the loading of Symbol Search.

All second order factor loadings were characterized as being high: FD = .90; VC = .86; PO = .85; PS = .62. Keith & Witta concluded from the high loadings of FD on the second order general factor that its name is a misnomer, failing to capture the higher order thinking skill its powerful loading actually suggests. This conclusion is consonant with that of Ownby and Carmin (1994).

Following these evaluations, Keith & Witta analyzed data using other theoretical models, including Kaufman’s three-factor model, the two-factor model related to Wechsler’s V/P dichotomy, and the one-factor model of Macmann and Barnett. Each resulted in a significant increase in the chi-square, suggesting the models produced worse fits than the four-factor model.

A three-stratum theory derived from Carroll’s work (1993b) was tested subsequently, yielding a fit that was slightly poorer than the four-factor model. Keith & Witta recommended further study of this model because although it is more

complex, it is also in some ways more parsimonious, perhaps offering a more powerful theoretical base for explaining what abilities are measured by the WISC-III. Keith & Witta echo Carroll's concern that the WISC-III in its present form does not include enough subtests with certain kinds of specificity (e.g., quantitative reasoning and memory) to test Carroll's model adequately.

Other Factor Analyses of the WISC-III with Non-Referred Samples

Donders (1996) constructed a short form of the WISC-III retaining two subtests from each factor. Vocabulary and Similarities were selected to represent VC. Block Design and Picture Completion were selected to represent PO. The two subtests for FD and PS were retained. Donders then subjected the 8 subtest short form to a confirmatory factor analysis using data from the performance of the WISC-III standardization sample. The purpose of his research was to determine which hypothesized factor model best explained the constructs being measured.

The four-factor model from the WISC-III manual (Wechsler, 1991) yielded the best fit, although it did not explain the data perfectly, as reflected in the statistically significant chi-square obtained. Donders compared the chi-square value from the four-factor short form model to the chi-square for the full-length form factor model. He concluded the model "is just about as imperfect for this short form as it is for the full-length WISC-III, and it is also the relatively best imperfect of all those models ... evaluated" (Donders, 1996, p.19).

Factor Analyses of Other National Samples

Roid, Prifitera, and Weiss (1993), employing an independent, nationally representative sample ($n = 1,118$), largely replicated the four-factor structure promulgated in the WISC-III manual (Wechsler, 1991) in a subsequent study. Subtests of the WISC-III and the Wechsler Individual Achievement Test (WIAT) were administered to this sample for purposes of conorming the WISC-III and WIAT. The sample of approximately 100 subjects at each age level was constructed employing the same stratification variables utilized in the construction of the WISC-III standardization sample from the 1988 Census Bureau. The stratified sample was racially balanced, including 75.6 % Caucasians, 12.4% African Americans, 9% Hispanics, and 2% other. Both exploratory and confirmatory analyses were conducted with this sample. Several extraction methods were utilized. Three, four and five-factor models were investigated.

From the maximum likelihood (ML) factor analyses a five-factor model was confirmed. This model included one factor involving a single variable, Comprehension, rendering it a very weak factor. The four-factor model proposed in the WISC-III manual (Wechsler, 1991) therefore was adopted instead. VC and PO accounted for 54% of the variance. FD, involving the same subtests as the WISC-III standardization sample, accounted for 6% of the variance. PS, again identical to the composition in the WISC-III standardization sample, accounted for 8% of the variance.

In the principal components analysis, VC and PO emerged as they had done in the ML analysis. Digit Span loaded alone on the fourth factor, however, and

Arithmetic split between VC and the fourth factor. The correlations that were generated in cross validation with the WISC-III standardization sample were .99, .98, .95, and .88, respectively, for VC, PO, FD, and PS. They provided evidence of factor structure reliability across these 2 samples.

Confirmatory analyses were conducted using 5 models; a one-factor model (all 13 subtests on one factor); Wechsler's V/P dichotomy two-factor model; Kaufman's three-factor model; the four-factor model proposed in the WISC-III manual (Wechsler, 1991); and a five-factor model. The latter involved a model examined by Wechsler (1991) and also by Woodcock (1990), with VC, PO, PS, Memory (i.e., Digit Span) and Numerical Ability (i.e., Arithmetic). Multiple goodness-of-fit indices were used. The five-factor model used in this study produced slightly better fit to the data. Roid et al. (1993), cautioned that there were many difficult statistical issues involved in accepting a singleton factor, however. They suggested further research would be needed to substantiate its validity. The emergence of the fifth factor in the form it took provided support, in fact, for Carroll's (1993b) conclusion that difficulties in establishing the factor structure of the WISC-III may be due to inadequate representation of subtests reflecting certain specificities.

The four-factor model was therefore accepted as being the most adequate with certain reservations. In support of the four-factor model, Roid et al (1993) noted the improvement in model fit (by chi-square differences) from the two and three-factor models to the four-factor model is much greater than from the four to the five-factor model. Additionally, the comparison of the four-factor model to the singleton

Comprehension factor did not in result in increase in model fit. Roid et al. concluded their cross validation study provided further evidence of the stability and replicability of the four-factor solution proposed in the WISC-III manual (Wechsler, 1991). They encourage clinicians to give Symbol Search and use the index scores. Because FD correlated .71 with the Mathematics Composite of the WIAT, Roid et al. suggest that FD measures more than short-term memory, possibly working memory that is relevant to math problem solving.

Roid and Worrall (1997) reported replication of the four-factor model using a stratified normative sample of 1,100 Canadian children, ages 6-16. Results of both exploratory and confirmatory factor analyses provide support for the factorial validity of the WISC-III four-factor model. Roid and Worrall (1997) note that the factor correlations of the Canadian sample closely matched those of the WISC-III standardization sample in the exploratory factor analysis. Results of the confirmatory factor analysis suggest that the chi-square values and other fit statistics reach a plateau at four factors. Roid & Worrall acknowledge the critiques of Carroll (1993), Sattler (1992), and Thorndike (1992) about the diminished eigenvalues of the third and fourth factors. Roid & Worrall suggest that “the present study shows remarkable robustness of these small factors to sampling fluctuations and cultural and historical differences across these two large, national normative samples” (Roid & Worrall, 1997, p.514).

Summary of Studies Involving Children with Learning Disabilities

Several studies have been conducted to date to assess the construct validity of the WISC-III in clinical samples that include children with learning disabilities.

Three clinical samples involving exclusively African American children were mentioned briefly in Chapter I under the heading “Racial/Ethnic Groups.” These will be reviewed initially in greater depth in this section.

The first study cited in the literature involving exclusively African American children was done by Slate and Jones (1995). They conducted a study of 58 African American students referred for special education including 19 with a specific learning disability, 22 with mental retardation, and 17 who did not meet eligibility requirements. Scores from 11 WISC-III subtests (Symbol Search was not included) were examined utilizing principal components factor analytic techniques. Varimax rotations were used “when appropriate” (Slate et al., 1995, p. 1070). Results were difficult to interpret with relation to factor models and the authors cautioned the small sample sizes, differing diagnoses, and limitations regarding geographic location raised concerns about the generalizability of the results (Slate & Jones, 1995).

Kush and Watkins (1997) investigate the factor structure of the WISC-III for a sample of 161 African American special education students who were undergoing triennial re-evaluations. The sample of exclusively Arizona urban/suburban students included 114 students classified as having learning disabilities, 10 classified as having emotional disabilities, 28 classified as having mild mental retardation, and 8 classified as having moderate mental retardation. The correlation matrix of the scores of these students on the 10 core subtests of the WISC-III was subjected to an exploratory factor analysis using maximum-likelihood extraction, followed by a varimax rotation procedure consistent with those conducted for the standardization sample of the WISC-III.

The first unrotated factor that emerged in Kush and Watkins sample accounted for 55% of the variance, demonstrating a coefficient of congruence of .99 when compared to the standardization sample. Kaiser, scree, and parallel analysis criteria were utilized to determine the number of factors to extract. Two factors, reflecting the traditional verbal/performance dimensions of the Wechsler Scales, were extracted. Following varimax rotations the coefficients of congruence for these two factors when compared to the standardization sample of the WISC-III were .99 for the Verbal factor and .98 for the Performance factor (Kush & Watkins, 1997).

Kush et al., (in press) investigated the construct validity of the WISC-III for 348 African American students referred for psychological evaluations. This sample, drawn from 10 states, included 206 students classified as having learning disabilities, 23 classified as having emotional disabilities, 25 classified as having mentally retardation, 2 classified as having speech/language impairments, 11 classified as having other health impairments, and 81 found ineligible to receive special education services.

Twelve subtests (excluding Mazes) of the WISC-III were subjected to maximum-likelihood exploratory factor analysis with Varimax rotation, followed by confirmatory factor analysis. The results of the maximum-likelihood exploratory factor analysis demonstrated the first unrotated factor accounted for 41% of the total variance, which yielded a coefficient of congruence of .98 with the standardization sample. Following rotation, the first two factors which emerged, accounting for 44% of the variance (versus 45% for the standardization sampled), conformed to the Verbal/Performance dimensions of the Wechsler Scales (both yielding coefficients of

congruence with the standardization sample of .99). A small third factor emerged, corresponding to the Processing Speed factor of the WISC-III standardization sample, involving strong loadings from Coding and Symbol Search and yielding a coefficient of congruence of .96 with the Processing Speed factor of the standardization sample. No evidence of the existence of the Freedom from Distractibility factor emerged. The results of the exploratory factor analysis therefore suggest a three-factor model.

The results of the confirmatory factor analysis, by contrast, suggested that the three-factor model was plausible (the four fit indices used suggested the models were equivalent). However, a structurally better fit to the data was obtained using the four-factor model, especially with reference to the improvement in chi-square values of the four-factor model over the two and three-factor models. Kush et al., (in press) suggest the discrepancies between the two analyses may be related to two psychometric considerations. The first involves difficulty in replicating factors when there are few observable variables per factor. The second consideration involves the fact that the use of varimax rotation can inflate small variables, removing from larger factors a portion of their appropriate variance (Kush et al., in press).

Six other studies of the WISC-III involving children with special needs are reported as well. The first of these investigated the factor structure of the WISC-III for a clinically mixed sample of 167 children with disabilities (ethnically unspecified), including children with learning disabilities, children with reading disabilities, and children with attention-deficit disorders (Wechsler, 1991). The age range of the sample was 6-16 years (median = 10 years). The sample was predominantly male (82%). Confirmatory factor analysis was used to test factor

models ranging from one to five factors. The four-factor model provided the best fit. However, Picture Arrangement had split loadings on three of the four factors.

This pattern was noted to be similar to the pattern of the 6-7 age group of the WISC-III standardization sample. The combined effects of small sample size, mixed samples, lack of explicitly identifying the proportion of each subsample that has each disability, lack of specification with regard to ethnicity, and failure to reproduce the covariance or correlation matrix and related fit statistics in total, raise concerns about the generalizability of results to other clinical samples of children with learning disabilities.

A study by Hishinuma and Yamakawa (1993) with an “at-risk” population of 78 children (including children with learning disabilities and children with giftedness, with only one percent representation of African American children) supported the WISC-III four-factor model. However, the researchers caution “the status of Picture Arrangement remains precarious. The loading of the subtest on PS strongly suggests this factor measures the constructs of visual-spatial discrimination, sequential processing, motor dexterity, and overall performance speed” (Hishinuma & Yamakawa, 1993, p. 103). The size of the sample is so small that generalizability of results obtained by Hishinuma and Yamakawa must be viewed with great caution.

Loderquist-Hansen and Barona (1994) compared the factor structure of the 12 subtests of the WISC-III standardization sample to samples of Hispanic (n = 120) and Non-Hispanic (n = 120) white children with learning disabilities. The age range of the children was 8 - 13. Both alpha factor analysis and canonical factor analysis were used, initially, to investigate the factor structure of the WISC-III in the

standardization sample and both samples of children with learning disabilities. Coefficients of congruence were then computed to compare the degree of factor similarity between the two clinical samples. The results of the factor analyses demonstrated a three-factor solution for the total sample and the two samples of children with learning disabilities when all twelve subtests were administered. For all of the groups the three factors consisted of Verbal Comprehension, Perceptual Organization, and Processing Speed. “The consistency across factor structure in alpha and canonical factor analysis suggest that these results are not method specific” (Loderquist-Hansen & Barona, p. 4). Coefficients of congruence between the three factors for both samples of children with disabilities were significant. The results did not support Wechsler’s fourth factor, Freedom from Distractibility in the standardization sample, the Hispanic sample with learning disabilities, and the sample of Non-Hispanic children with learning disabilities.

Bell (1994) investigated the factor structure of the WISC-III for a sample of 246 children (169 boys; 177 girls) with learning disabilities residing in Arizona. The subjects ranged in age from 6 - 16 (average age = 10.27 years). The ethnic composition of the sample was 67% Caucasian, 19% Hispanic, 13% African American, and 1% Native American. Confirmatory factor analyses were conducted on the covariance matrix of the sample for four models, a one-factor model, a two-factor model, a three-factor model, and a four-factor model. Fit statistics were used to assess model fit including the chi-square, standardized root mean square residual (RMSR), goodness of fit index (GFI), and adjusted goodness-of-fit index (AGFI). Model modification involving adding error covariance was allowed under two

conditions. First, when pairs of subtests had large, positive residuals and were included in the same factor, error covariance was added. Second, when pairs of subtests had large negative residuals and the two subtests were not included in the same factor, error covariance was added.

The GFI values were acceptable for Models 2, 3, and 4. The AGFI values were acceptable for Models 3 and 4. The RMSR was acceptable also for Models 3 and 4. Each model incorporating more factors reflected successive improvement in fit over the preceding models with fewer factors with respect to the chi-square statistic. Bell concluded the four-factor model provided the best fit to the data. Additionally, the modified version of Model 4 resulted in the lowest chi-square value in relation to degrees of freedom with respect to all unmodified and modified versions of models. Given the relatively small sample size, Bell cautioned the model modifications might not generalize to other samples. Additionally, the choice of the GFI and AGFI as additional fit indices may be problematic in that they have been shown to be biased in samples under 250 (Marsh et al., 1988).

Kush (1996) investigated the factor structure of the WISC-III for a sample of 327 children (228 boys; 99 girls) with learning disabilities attending schools in the Southwest. The ethnicity of the sample was reported as follows: 158 Caucasian; 54 Mexican-American; 36 African American; 78 Native-American; 1 Asian/Pacific Island. The socioeconomic status of the children was characterized as including low-middle and middle-class strata. The correlation matrix from 12 of the WISC-III subtests (excluding Mazes) was subjected to maximum-likelihood exploratory factor analyses using both oblique and varimax rotations.

Results from the analyses using oblique and varimax rotations were very similar. In both cases, a three-factor solution emerged, with the verbal tests all loading substantially on a verbal factor except Digit Span. All of the performance tests except Symbol Search and Coding loaded substantially on a performance factor. Symbol Search and Coding formed a third factor that accounted for only 5% of the variance. Digit Span had low positive loadings on all three factors.

Konold, Kush, and Canivez (1997) investigated the factor structure of the WISC-III for three samples of children with disabilities. The first sample, a mixed sample, consisted of 229 children (69% boys; 31% girls). About 81% of that sample were classified as having learning disabilities. About 8% were classified as having mild mental retardation. About 7% were classified as having emotional disabilities. About 3% were classified as having speech/language impairments. Less than 1% were classified as other health impaired. Less than 1% were classified as moderately mentally retarded. The ethnicity of the sample was reported as follows: Caucasian = 51%; African American = 10%; Hispanic = 33%; Native American = 4%; and Asian = 2%. The second and third samples were described previously. The second sample was acquired from Bell (1994). The third sample was acquired from Loderquist-Hansen and Barona (1994).

Covariance matrices of the three samples were subjected to maximum-likelihood confirmatory factor analyses using LISREL 8 (Joreskog and Sorbom, 1993) across five models. The models were analogous to those reported in the WISC-III manual (Wechsler, 1991). Fit statistics in addition to the chi-square included the adjusted goodness-of-fit index (AGFI) and the standardized root mean

squared residual (RMSR). Model improvement statistics reported included the likelihood ratio chi-square statistic, the Tucker Lewis index (TLI) and comparative fit index (CFI). The TLI was used to evaluate both an independence model and a one-factor model. The one-factor model served as a baseline against which the two-through five-factor models were contrasted.

The four-factor solution provided the best fit with respect to the likelihood ratio chi-square test and the TLI. It provided better estimates than one-, two-, and three-factor models regarding the goodness-of-fit indices. It matched or exceeded estimates for those indices obtained by the five-factor model (Konold et al., 1997).

Konold et al. (1997) cautioned that the heterogeneity of classifications in the samples is a limitation with respect to generalizing results to other samples of children with learning disabilities. Additionally, they noted it was not possible to obtain the criteria used to classify the children with learning disabilities to determine if they were uniform. Additionally, specific information about the individual types of learning disabilities that each child in each sample had was unavailable, making it impossible to determine how heterogeneous these samples were. Lastly, the samples all were drawn from the southwestern region of the U.S. only.

Summary of Studies of Patterns of Abilities

If an intelligence test reflects a similar structure of abilities consistently across racial and clinical groups, it is possible to compare the profiles of scores obtained by different subgroups to determine if there are variations in the patterns of scores between groups. A considerable body of research has been amassed on tests of abilities which suggests there are differences in the profiles of abilities among ethnic

groups and some clinical populations (Suzuki, 1992; Vernon et al., 1988; Reynolds & Kaufman, 1985; Reynolds & Jensen, 1983). These differences can be evaluated by comparing individual subtest scores or performance on a cluster of subtests which are thought to be reflective of specific mental processes (e.g., memory, verbal fluency, numerical ability, etc).

Three studies involving levels and patterns of performance in the WISC-III standardization sample, one study involving the WISC-III/WIAT linking sample, a referred sample of children in a special education program, and one study involving a sample of children identified as having learning disabilities have been reported to date. The study involving the standardization sample and the linking sample provide important baseline data about the patterns which children normally show with regional variations in ability and achievement. The degree of variation in the WISC-III and WIAT standardization samples serves as a marker for comparison with clinical populations in order to make determinations whether they display unusual patterns. If unique profiles or patterns can be identified in clinical samples which distinguish them from the standardization samples they can be investigated with respect to both the diagnostic and treatment utility of the subtest scores. If unique patterns cannot be determined the practice of subtest analysis will be challenged. Table 5 contains a summary of the research generated with respect to the WISC-III standardization sample, the WIAT linking sample, the referred sample, and a sample of children with learning disabilities. Detailed information regarding these studies is in Appendix B, entitled "Studies of Patterns of Abilities of Samples Using the WISC-III Subtest Scores."

Differences in general ability were identified consistently as the primary distinguishing characteristic in all four studies using data from the WISC-III standardization sample. Full scale IQ was identified as the most discriminating factor in the WISC-III/WIAT study (T. J. Ward, S. B. Ward, Glutting & Hatt, 1999) involving children with learning disabilities. Using the multivariate methodology provided by Glutting et al. (1994) to examine the ability and achievement score of 201 children with learning disabilities, T. J. Ward et al. (1999), determined 70 % of the children in their sample presented patterns matching a core profile identified in the Glutting et al. (1994) study of children without disabilities. The incidence of occurrence of unusual profiles in the sample of children with learning disabilities (Ward et al., 1999) did not differ greatly from the occurrence of those profiles in the standardization sample with Glutting et al. (1994) study. Glutting et al. (1994) had concluded the pattern of performance of children in the standardization sample reflects sufficient variability to warrant approaching the clinical practice of subtest analysis in generating hypotheses concerning children's learning problems with great caution. Ward et al. (1999) provided further evidence "that many children identified with LD actually demonstrated profile of performance that children normally show" (Ward et al., 1999, p 640). The conclusions reached by these authors are consistent with those reached by Kavale and Forness (1984) and Watkins and Kush (1994) in studies involving the WISC-R.

Table 5

Studies of Patterns of Abilities Generated from WISC-III Scores

Author(s) & Date	Sample ^a	Subtests	Methodology ^b
Glutting, McDermott, Prifitera, & McGrath, 1994.	1	12 excluding Mazes	1
S. B. Ward, T. J. Ward, Hatt, Young, & Mollner, 1995.	2	Selected Subtests associated with profiles	2
Donders, 1996.	3	12 excluding Mazes	3
Glutting, McDermott, & Konold, 1997.	3	13	4
Konold, Glutting, McDermott, Kush, &Watkins, 1999.	3	10 core subtests	4
T. J. Ward, S. B. Ward, Glutting & Hatt, 1999.	4	4 index scores from 12 subtests	1

^a1 = WIAT linking sample; 2 = Sample of 719 children receiving special education services; 3 = WISC-III standardization sample; 4 = Sample of 201 children with learning disabilities. ^b1 = Multivariate cluster analysis; 2 = rate of occurrence of 3 profile associated with disabilities; 3 = SAS and Fastclus Cluster analysis; 4 = nonlinear multivariate methodology.

A Review of the Literature on Test Bias as it Relates to Validity

It is evident in the federal regulations that test bias is a crucial dimension that must be addressed in selecting tests for use in determining if students are eligible for special education programs. Test bias specifically, by definition, is “systematic error in the estimation of some true value for a group of individuals” (Reynolds, 1982, p. 186). “Claims of bias must be based on objective evidence, just as claims of validity and reliability must be substantiated by empirical investigation” (Reynolds, 1982, p. 185). It was not the purpose of this study to focus comprehensively on issues involving test bias in relation to intelligence tests. This study of the structure of abilities of African American school children is relevant primarily to a discussion of construct validity with practical implications for test use in schools. However, it was important in this chapter to review briefly empirical research pertaining to aspects of test bias in relation to African American school children because of empirical support it provides for the use of cognitive ability measures with native-born English-speaking racial/ethnic subgroups.

Drawing upon research of Clarizio (1978), Reschly (1980), Vandivier and Vandivier (1979) and Wright and Isenstein (1977), Reynolds, Lowe, and Saenz (1999) classified dimensions of empirical test bias research into seven categories: inappropriate content; inappropriate standardization samples; differential predictive validity; examiners’ and language bias; inequitable social consequences; measurement of different constructs; and qualitatively distinct minority and majority aptitude and personality. Extensive research has been conducted on these dimensions in relation to intelligence tests since the 1970’s, yielding complex results (DiGangi &

Faykus, 1993; Reynolds et al., 1999). “Psychometrists have been working diligently, but all of the answers are not yet in” (Reynolds et al., 1999, p.584).

Content, Standardization Samples and Predictive Validity

Significant changes were made in the development of intelligence tests during the last three decades in relation to content, standardization samples and predictive validity. Prior to the 1970's the most commonly utilized intelligence test, the Wechsler Intelligence Scale for Children (Mardell-Czudnowski & Burke, 1991; Reynolds, 1982) did not even include minority children in its standardization sample. Substantial changes in intelligence tests developed more recently have resulted in better sampling procedures. These include incorporating proportionate numbers of minority children in the standardization samples (Wechsler, 1991). Item content is carefully reviewed to see if items fit the behavioral domain which is being assessed (Anastasi & Urbina, 1997), and if they are biased against subgroups. Additionally, statistical analysis of the possibility of differential prediction across both ethnic and gender groups is conducted (Wechsler, 1991). Brown, Reynolds, and Whitaker (1999) state, “current frequently used mental tests are perhaps the most carefully constructed, standardized, and evaluated of all psychological tests” (p. 231). Access to large, stratified samples is an advantage afforded to major test developers (Brown, et. al., 1999) including the Psychological Corporation which developed the WISC-III. Additionally, item response theory (IRT) based methods for detecting item bias in very large samples (which are evaluated using computer-processing capabilities available to major test publishers) are used currently by all developers of major tests of intelligence. Extensive research on intelligence tests and other ability tests across

various racial groupings utilizing a diverse set of criterion measures have not revealed differential prediction (Brody, 1985; Brown et al., 1999; Jensen, 1998; Reynolds, 1982; Roid, Prifitera & Weiss, 1993; Sattler, 1992).

Examiners' and Language Bias

Questions related to the influence of examiner's race on examinee's performance have produced inconclusive results (Graziano, Varga, & Levy, 1982; Lutey & Copeland, 1982). Reviews of the literature on examiner bias (Clarizio, 1978; Jensen, 1980; Pryzwansky, Nicholson, & Uhl, 1974; Shuey, 1966) have concluded that using a Caucasian examiner does not impact negatively upon the validity of test results for ethnic minority children. Moore and Retish (1974) examined the effect of race of the examiner on the performance of 42 African American children on the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; 1967). The effect of the race of the examiner was significant for FSIQ with African American examiners eliciting higher mean scores ($M = 93.21$) than Caucasian examiners ($M = 87.74$). Savage (1972) reported results of a study where black children scored higher on a Block Design task but not Digit Span with black examiners than with white examiners.

In reviewing 29 studies which attempted to address this problem, Graziano et al., (1982) concluded there was no definitive, consistent evidence that the race of the examiner elicited differential performance among minority children. Graziano et al., reported the confounds in studies presenting contradictory evidence are related to inadequate designs and incomplete sampling. "Furthermore, lack of appropriate control groups, instances of unbalanced treatment presentations, possible nonrandom

assignment of examiners to treatment conditions, and inappropriate data analysis all make unequivocal interpretation of the literature difficult” (Graziano et al., 1982, p. 491).

The question of a differential effect when utilizing black dialect versus standardized English in administering test directions has yielded mixed results (Reynolds, 1982). “However, where the purpose of the test is to assess the knowledge of standard English concepts and vocabulary, the differential effect of wording is not an appropriate question” (Reynolds, 1982, p 182). Jencks (1972) concluded that no evidence exists that African American children are at a more significant disadvantage on verbal tests than other groups. Quay (1974) and Crown (1970) found black children did not do better on tests when presented directions in black dialect than controls tested utilizing directions in English.

Inequitable Social Consequences

Has the use of intelligence tests for the purpose of classification of children as being eligible for special education services resulted in inequitable social consequences for some minorities? Values particularly enter the evidence-gathering process with respect to this question. They are a legitimate consideration for study in examining bias in test use. Cole and Moss (1989), Cronbach (1976), and Messick (1980,1981) all incorporate values and evidence about test consequences in their guidelines for investigation of validity. Cole and Moss (1989) note it is incumbent upon investigators who evaluate bias in test use to examine both values and evidence concerning outcomes. Whether one considers study of the worthiness of the purpose of test use to be integral to the study of validity (e.g., Messick, 1975) or an extra-

validity issue (Cole & Moss, 1989), explicit articulation of value judgments by investigators is essential.

When interested parties in a testing situation have different purposes and these remain unarticulated, debates might never reveal the fundamental reasons for differences of opinion. It is the failure to articulate underlying purposes and values that has led to many inconclusive debates about bias in testing. (Cole & Moss, 1989, p. 215).

The issue of inequitable social consequences involves in part determining whether being identified or labeled and provided with a special program places those individuals at a disadvantage educationally and vocationally. On the one hand, some minority spokespeople have taken the position that placement in special education programs has been racially discriminatory. At the annual meeting of the American Psychological Association in Washington, D.C., in 1969, the Association of Black Psychologists and subsequently other minority spokespersons, including R. L. Williams (1971, 1970) explicitly suggested that the use of intelligence tests and other standardized measures have resulted in over classification of African American children.

Controversy related to these allegations has been so intense that it has reached the federal courts where legal remedies have been sought to address this issue. Diverse perspectives are reflected initially in rulings by federal judges in two different states. In *Parents in Action on Special education et al. v. Joseph P Hannon et al.*, No. 74 C 3586 (1980), in Chicago, Judge John F. Grady ruled that although some items on intelligence tests may be interpreted as biased, their overall effect on the decisions made regarding African American children would be minimal (i.e., not discriminatory against these children) (Sattler, 1981). In *Larry P. et al. v. Wilson*

Riles et al., No. C-71-2270 RFP (1979), in Northern California, Judge Robert F. Peckham initially took the position that intelligence tests used for the purpose of assigning students to special education classes was discriminatory and banned the use of intelligence tests in the state of California for this purpose (Sattler, 1981).

Sattler (1981) reviewed the perspectives of the two federal judges through an examination of their final decision opinion statements. The purpose of this review was to shed light on the values and judgements of the two judges that informed their diametrically opposed decisions. An examination of statements contrasting their opinions revealed a perception on the part of Judge Peckham that placement in special education classes was a bias issue for him in and of itself. He perceived classes for educable mentally retarded children as providing an inadequate curriculum for children whom he described as being incapable of learning. Additionally, his view of intelligence tests was that the entire placement process revolved around the determination of IQ. Judge Grady, by contrast, conceived of it as being part of “a much broader perspective of the decision-making process” (Sattler, 1981, p. 361). More serious than this issue is Judge Peckham’s stated belief that intelligence tests were constructed by psychologists who held racist views and “intentionally devised tests that placed blacks in an inferior position” (Sattler, 1981, p. 363). Additionally, he believed bias operated against African American children with respect to use of standard English, effectively accepting selective testimony and disregarding evidence to the contrary. More recently, Judge Peckham’s ruling has been overturned, bringing consistency to federal courts (Reynolds, 1995).

Psychologists who utilize intelligence tests and other ability measures to assist in determining eligibility of children for special education services defend this practice vigorously by arguing that cognitive measures generate objective data that document educational deficits, evidence which they value highly. They assert test data can be utilized to obtain needed services for underachieving children who would otherwise be vulnerable to academic failure without intensive intervention (Oakland & Parmelee, 1985; Sattler, 1982). Reschly (1981) underscored the fact that special education classification and placement lead to expenditure of substantially more money on the education of children with disabilities than remedial and regular education. Oakland and Parmelee argued, "The use of standardized tests represents important components of a nondiscriminatory assessment program" (p. 716). Reschly (1981) concluded, "I.Q. test use protects many students of all races, social statuses, and genders from erroneous and inappropriate classification" (p. 1097), a negative side effect which he believes would be more likely to occur if results of standardized assessment instruments were not employed.

Empirical evidence suggests African American children are less likely to be recommended for special education services particularly special education classes than Caucasian children who are presenting similar cognitive issues (Frame, 1979; Matuszek & Oakland, 1979; Meyers, MacMillan & Yoshida, 1978; Reynolds, 1982). McLeskey, Waldron, & Wornhoff (1990) suggest a key factor with respect to children suspected of having a learning disability is the choice of method used to determine whether a severe discrepancy exists between the measure of intelligence and the measure of achievement. McLeskey et al., reported most agencies use standard score

difference procedures, expectancy formulae, or grade deviation procedures and/or cutoff scores rather than regression based methods. In their analysis, McLeskey et al., determined methods other than regression resulted in a disproportionate underrepresentation of African American children meeting eligibility criteria.

Reynolds and Kaiser (1990) in reviewing research on labeling effects noted that psychologists tended to rate the intelligence of African American children higher than Caucasians utilizing the WISC. They concluded that “Black and low SES children are less likely to be recommended for special education placement than their white or high SES peers with similar cognitive, behavioral, and emotional characteristics” (Reynolds & Kaiser, 1990, p. 618).

Evidence of Construct Bias

Do intelligence tests measure different attributes for minority children than the standardization sample of the intelligence test? Reynolds stated “bias exists in regard to construct validity when a test is shown to measure different hypothetical traits (psychological constructs) for one group than another or to measure the same trait but with differing degrees of accuracy” (Reynolds, 1995, p.559). Jensen (1998) reported the most extensively researched groups with respect to the question are African Americans of South African ancestry and Caucasians of European ancestry. After exhaustively reviewing research generated from 1970-1990 regarding construct bias, Reynolds and Kaiser (1990) concluded, “No consistent evidence of bias in construct validity has been found” (p. 638). More recently the comprehensive “Report of the Task Force Established by the American Psychological Association” (Neisser et al., 1996) reviewed critical issues about what is known about intelligence. The authors

acknowledged in their report that many different perspectives exist about what intelligence is and how to measure it. However, in agreement with Reynolds and Kaiser, they concluded no evidence of bias in construct validity had been found. Neisser et al., reported, "The differential between the mean intelligence test score of Blacks and Whites (about one standard deviation, although it may be diminishing) does not result from any obvious biases in test construction" (Neisser et al., 1996, p. 97). Jensen (1998) affirmed this conclusion across all culturally diverse ethnic groups in the United States as did Brown, Reynolds, and Whitaker (1999).

A narrow focus in the investigation of construct validity addresses the question of whether tests that are standardized on a majority Caucasian sample or an exclusively Caucasian sample tend to be systematically biased against ethnic groups that have smaller representation or no representation in the standardization sample of the test. Fan, Willson, and Kapes (1994, 1996) investigated these questions in an experiment. Two test construction models representing these conditions (i.e., 100% representation of one ethnic group, and differential representation of Caucasian, African American, Hispanic, and Asian groups) were studied. Fan et al., (1996), concluded there was "no systematic bias against the group(s) with smaller or no representation in the test construction samples" (p. 365). Their findings lent support to current psychometric procedures which are widely employed in relation to sampling and item selection (Fan et al., 1996).

Qualitatively Distinct Minority and Majority Aptitude and Personality

Are racial differences in cognitive ability based upon "deficient conceptualizations of culture?" (Reynolds et al., 1999, p. 584). Reynolds et al. (1999)

exhaustively reviewed the literature in this area in relation to the claims of Helms (1992) that cognitive ability in ethnic minorities is different from that of members of the dominant culture and that insufficient research has been conducted to date to address the culture issue (Reynolds et al., 1999). The conclusion they reached is that

... research shows that IQ tests appear to measure the same psychological characteristics in blacks ... and Anglo-Americans (Rowe, 1994). ... If Helms (1992) were calling for continued research into these areas such a call would have our support, but to dismiss the many studies that are addressing these issues is improper (Reynolds et al., 1999, p. 584).

Environmental Issues Involving Assessing the Intelligence of Ethnic Minority Children Living in Poverty

Childhood poverty in the United States has been on the increase since 1989, resulting in an estimated 22% of American children living in poverty in 1994 (U.S. Bureau of Census, 1996). McLoyd (1998) reports American children under the age of 6 are at even greater risk of being poor than school age children. American children are worse off with respect to poverty than their counterparts were three generations ago (McLoyd, 1998).

Although African American children account for about 15% of the child population in the United States (U.S. Bureau of Census, 1994), they are two to three times more likely to live in childhood poverty than non-Latino White children (U.S. Bureau of Census, 1996). The poverty rate among African American children under the age of 18 in 1988, for example, was 44.2% in contrast to the poverty rate of 14.6% for Caucasian children (U.S. Bureau of Census, 1988). The rate of poverty has increased more rapidly for African American children than Caucasian children in recent decades. For example, between 1979-1985 the rate of poverty for African

American children increased from 36% to 41% while the increase for Caucasian children was from 12% to 13% (Salkind, 1994). Additionally, African American children are more likely than non-Latino White children to experience persistent poverty, and under those circumstances to live in areas where there are greater concentrations of poverty (Duncan & Rodgers, 1988; Jargowsky, 1994). McLoyd (1998) reports greater geographical concentration in inner-city neighborhoods of poverty among poor African Americans than poor Whites.

The effect of the context of poverty on child development in general has been the subject of recent study based upon Bronfenbrenner's conceptual work (1986). J. J. Wilson's (1987) scholarly historical analysis of the changes that led to inner-city concentrations of poverty among African Americans, is a second important source of information that has heightened understanding of relevant research variables. It is noteworthy for its focus on children, especially with regard to the possible effects of poverty on aspirations, attitudes and opportunities for support and mentoring from caregivers and other significant adults.

Recent contributions from the field of developmental science have enabled researchers to reconceptualize and redirect their study of the effects of poverty upon African American children and families. McLoyd (1998) summarized the changes brought about by these conceptual and methodological advances, which involve a focus on the timing and duration of poverty and the expansion of the study of the ecology of poverty to take into account "the multiple levels of proximity to children" (McLoyd, 1998, p. 188). Additionally, the definition of economic well being has been expanded beyond the cash income marker used in federal government

determinations. It now incorporates measures of parental occupation, parental education, family income, prestige, and power (McLoyd, 1998). Analyses of the processes by which one becomes, and is maintained, as socioeconomically disadvantaged is another component in this research thrust (McLoyd, 1998). Additionally, the methodology of carefully constructed longitudinal studies adds an important dimension (McLoyd, 1998).

McLoyd (1998) reviewed risk factors involving caregivers. She reported African Americans have achieved lower levels of education, are employed in jobs which are less stable, and are employed in jobs which have lower rates of reemployment following layoffs. Additionally, they are less likely to live in homes in suburban and other nonmetropolitan areas where greater growth is occurring in entry-level jobs. Females head more homes, as well, introducing another factor related to poverty status (Eggenbeen & Lichter, 1991). Evidence of harsh and inconsistent parenting by caregivers is reported at elevated levels as well as their exposure to acute and long-term stressors (McLoyd). Age of the mother impacts on intelligence independent of maternal education and SES (Broman, Nichols & Kennedy, 1975).

Environmental variables studied that impact upon intelligence of children include issues related to neonatal and infant mortality such as low birthweight and prematurity. Duncan (1994) studied the longitudinal data from the Infant Health and Development Program of low birthweight, premature infants (the sample was 55% African American). He concluded family income and poverty status significantly predicted IQ in 5-year-old children even after the effects of maternal education, family structure, ethnicity, and SES were accounted for in this study. Additionally,

the poverty status of the 3-year-olds predicted IQ at age 5. Mean IQ scores of children in this study were three quarters of a standard deviation lower than those of nonpoor children.

Jensen (1998) reported a disproportionate number of babies who are either premature or of low birthweight continue to be born to African American mothers. These babies are at greater risk for infant mortality than Caucasian babies. This particular risk factor exists across all SES levels for African American women. Higher rates of lack of access to health care, both medically and with respect to dental hygiene and care, including immunizations, are significant issues for many African American adults living in poverty and their children (McLoyd, 1998). These are associated with higher rates of sudden infant death syndrome (SIDS) and AIDS. Jensen cited “racial stressors” as a contributory factor in this health issue for African American women.

Children living in poverty are at greater risk to receive inadequate nutrition. Eysenck (1991) believed nutritional deficiencies are a significant cause of decrements in intelligence in children living in poverty. He included in these nutritional factors, dosages of vitamins and minerals that optimize cognitive growth and functioning. Eysenck’s research suggested these nutritional supplements could increase IQ gains on tests measuring fluid intelligence. He believed the diets of African American children are likely to be deficient in these critical nutritional elements.

A recent report by Lucas et al. (1992), links intelligence to use of breast feeding in preterm babies. In that study from birth into middle childhood (age 7.5) there was a significant linear dose-response relationship between the intelligence

quotient of the child and the amount of breast milk the child received. African American babies, although at greater risk for low birth weight, are less likely to be breast-fed (Jensen, 1998).

Other environmental variables impacting negatively upon intelligence in children living in poverty include greater risk for exposure to environmental teratogens including lead, alcohol and drugs, both illegal and legal, and the effects of tobacco smoking (Klerman & Parker, 1991; McLoyd, 1998; McLoyd, 1990; Salkind, 1994; Zill & Schoenborn, 1990). Children living in poverty are at greater risk for exposure to drugs than the general populace (Jensen, 1998).

Children living in poverty are at greater risk for inadequate cognitive stimulation in the home involving language and other preacademic skill development (Klerman & Parker, 1991). They also are less likely to be impacted by macroenvironmental influences including preschool or cognitively oriented day-care facilities. Early intervention programs have been created to target young African American children in this at-risk category and to attempt to impact upon their cognitive development. The most widely utilized of these is Head Start, a federal preschool intervention program, which has been in continuous existence across the United States since its inception in 1964. The second most publicized of these was the Milwaukee Project, inaugurated in the 1960's with follow up through 1981. Outcome measures have not reflected hoped for substantial long-term gains.

The Abecedarian Early Intervention Project, a less well known study, begun in 1972, at the University of North Carolina, produced a more powerful effect size. It targeted children at risk for mild mental retardation, providing medical, nutritional,

and cognitive/educational interventions over 5 years, beginning in early infancy. Ninety-eight percent of the recipients in this program were African American. Jensen (1998) characterized it methodologically as being conducted “in a model fashion as a scientific experiment” (p. 343).

At ages 8, 12, and 18, participants demonstrated maintenance of an IQ advantage of five points, and comparable advantages regarding scholastic achievement. Approximately half as many children who received the comprehensive program were retained as the control group. The control group received medical, nutritional, and family support services provided by social workers but not the cognitive curriculum. Additionally, the percentage of children testing with IQ < 85 averaged only 28% in the group receiving the comprehensive program as compared to 44% in the control group. This outcome measure was considered by Jensen to provide evidence that certain types of early intervention can impact positively on the cognitive development of young, “at risk” ethnic minority children.

Living in poverty increases a child’s chances of being placed in a special education program during his/her school years (McLoyd, 1998).

It seems dubious to assume that conditions believed to be related to learning disabilities and emotional disorders, such as, poor prenatal care, malnutrition, limited developmental and educational opportunities, are equally distributed across racial groups. Minorities often experience unique stresses, such as, prejudice and discrimination, lower socioeconomic status, problems of acculturation, which would be expected to cause more distress symptoms among minorities than among other racial groups (Griffith, 1994, p. 5).

Kamphaus (1993) reported that one possible effect of these environmental risk factors on low SES students over time is that they may produce intelligence test score decrements in relation to the normative mean (Anastasi, 1988; Saco-Pollitt, Pollitt, &

Greenfield, 1985). Jensen (1998) suggested this empirical phenomenon, commonly referred to as “the theory of cumulative deficit” (p. 495) gained acceptance in the 1960s as a theory explaining the “age-related decrement in IQ and achievement” (p. 495) in African American children in relation to norms for Caucasian children. According to the theory, early environmental disadvantage impacting negatively upon initial learning depresses both IQ and achievement over time, compounding with age, and hindering learning at later stages of development (Jensen, 1998). Research supporting this phenomenon includes the work of Sherman and Key (1932) with mountain children in Appalachia living in poverty, the research of Misra (1983) regarding low SES children in India, Jensen's (1977) study of low SES African American children, and Cox's (1983) study of low SES children in England and Wales.

Empirical Studies Involving Ethnic Minorities

Artiles, Trent, and Kuan (1997) reported a paucity of empirical studies have been published related to the classification of ethnic minorities in special education. They reviewed 2,378 empirical articles published in four major learning disability and special education journals between 1972-1994. The proportion of articles published with respect to ethnic minority students did not exceed the range of 6-8% during that period (Artiles et al., 1997). Artiles et al. criticized the methodological characteristics of that research as being generally weak. They suggested that studies should be conducted in “distinct settings with single disability and ethnic groups” (p. 90) in order to disaggregate the effects of ethnicity and disability status (Artiles et al., 1997). If local agencies would rely upon empirically sound research in decision-making in

relation to ethnic minority children, there is hope the consistency and accuracy with which children are classified and provided with efficacious services would be strengthened.

Methodological Issues in Conducting Research on Children with Learning Disabilities

Durrant (1994) explored the quality of learning disabilities (LD) research from 1988-1990 in 10 major journals, by comparing it to two standards. The first standard is the criterion developed by Torgesen and Dice in 1980 to evaluate LD studies. These threefold criteria included specification of the most descriptive variables, consistency in operationally defining learning disabilities, and strengthening of methodology. The second criterion involved a set of standards for publication of research adopted by the Council for Learning Disabilities in 1984. These were strongly influenced by recommendations of Torgesen and Dice. The standards encourage explicit reporting of “gender, age, race, socioeconomic status, IQ, achievement, type of educational placement, and school districts’ placement criteria” (Durrant, 1994, p. 25). Additionally, Durrant wanted to determine if changes were being made in the type of LD research conducted.

Durrant, using the same criteria Torgesen and Dice had used in 1980, set the following standards for studies he would review. First, the focus of the study must be children with learning disabilities. Second, the study must report empirical data. Third, it must attempt to report information concerning the psychological characteristics of the sample or evaluate the effectiveness of nonmedical interventions. Fourth, children with attention deficit disorders would not be the

primary focus of a study to be included. Fifth, the study must not be designed narrowly to assess the relationship between two test instruments.

Two hundred eight articles that met the above criteria were reviewed and results of the overall analysis were compared to the results obtained by Torgesen and Dice in 1980 to assess whether progress had been made in strengthening LD research overall. The number of published articles meeting the criteria increased approximately 50%. Two journals not reviewed previously, *Learning Disability Quarterly*, and *Learning Disabilities Research*, accounted for 56 additional articles. About 52% of all eligible articles were published in a single journal, the *Journal of Learning Disabilities*.

Content area foci were examined. Durrant reported a 100% increase in the proportion of studies involving intervention effects. Literacy dependent variables were most commonly used in these studies, followed by social and behavioral characteristics. Few developmental or longitudinal studies were noted, causing Durrant to conclude they were not being emphasized in LD research.

Durrant (1994) viewed the frequency of reporting major subject descriptive variables next. Gender was generally reported, although the 4:1 male/female ratio generally accepted as the rate of occurrence by gender was not matched in the majority of the studies reviewed. Age information was not uniformly included. Although there was an increase in studies involving adolescents, from 14%-36%, only 7% of the studies exclusively involved teens (Durrant, 1994).

Durrant's (1994) findings were congruent with the findings of Artiles et al. (1997) with regard to ethnicity in the LD literature. Failure to report ethnicity, use of

solely Caucasian samples with generalization of conclusions across other ethnic groups, and lack of visibility in ethnic minorities in samples were some of the findings reported by Durrant. He noted that 36% of the studies did not include African American children.

Although socioeconomic status (SES) is an important demographic variable, 59% of the studies reviewed by Durrant did not report it with respect to their samples, and an additional 30% did not use precise terms and measurements in specifying it. Equally imprecise reports appear with regard to information about the intellectual levels of children who participated in the studies. Significant inconsistency was noted in which scores were reported when scores were included (i.e., FSIQ, VIQ, PIQ, prorated, and unspecified). Highly heterogeneous IQ ranges were reported (minimum 40-120, and maximum 84-144). "In only 51% of the studies reporting IQ ranges did subjects' IQ scores fall between 80 and 120 exclusively" (Durrant, 1994, p.28). The most commonly used general tests of intelligence were the Wechsler Intelligence Scale for Children-Revised and the Wechsler Adult Intelligence Scale (used in 85.8% of the studies), followed by the Stanford-Binet (9%), and the Kaufman Assessment Battery for Children (5%).

Achievement levels of children, albeit limited, were reported in 65% of the studies. Achievement area tests were unspecified in many cases, rendering it impossible for the reader to be clear about the specific nature of the disability. In studies reporting content area measured, 96% evaluated some aspect of reading, 40% math achievement, 14% spelling achievement, 8% written language, 8% language skills, and 7% overall achievement. Of the 47 achievement tests reported, versions of

the Wide Range Achievement Test (21%), the Woodcock-Johnson Psycho-Educational Battery (13%), the Woodcock Reading Mastery Test (10%), and the Peabody Individual Achievement Test (8%) were the most commonly reported, revealing a high degree of heterogeneity in measures.

More than a third of studies provided no information about school placement. When placement was specified, most students were reported to be receiving resource room instruction. About 30% were in self-contained classrooms, 16 % in regular classrooms, and 7% in special schools.

Diverse procedures were reported with regard to definitions of LD. Approximately 70% of children were selected on the basis of prior identification using federal, state, or local agency criteria, these being unspecified in 33% of the cases. The identification of a severe discrepancy between ability and achievement was specified in 98% of the cases but how that discrepancy was quantified was reported in only 58% of the cases. Only 2% of these studies reported cutoff scores. In about 10% of the studies, the selection of children was based on research criteria, half of these adhering to discrepancy formulae. About 80% of the discrepancy criteria were operationally defined. A cutoff score was employed in 42% of the studies using a research definition of LD.

After reviewing methodological practices that might limit interpretability of findings, Durrant made several observations. First, he noted issues related to the heterogeneity of LD samples were not acknowledged in the studies he reviewed. Second, researchers seemed to be employing dependent measures not used in previous research less frequently than Torgesen and Dice had noted. Third,

discussion of validity occurred with somewhat greater frequency in currently reviewed studies. Fourth, the incidence of manipulation of experimental variables decreased in current studies reviewed. Fifth, few attempts were made to match target and control groups on behavioral measures. Durrant cautioned that although ability matching raises many serious methodological issues, the failure of researchers to attend to the importance of intelligence as a potentially confounding variable cannot be ignored with respect to research where achievement dependent measures are employed. He concluded, "Overall, reporting and design of LD research still falls far short of fully addressing the recommendations made by Torgesen and Dice (1980) and CLD (Smith et al., 1984)" (Durrant, 1994, p. 31).

Summary

The present study is a theoretical study focusing on the psychometric properties of the intelligence test most commonly used with children and adolescents, the WISC-III. Question investigated in this study involved the equivalence of constructs measured by the WISC-III in two conditions (i.e., Caucasian children with out disabilities and with disabilities, and African American children without disabilities and with learning disabilities). The review of the literature concentrated on summarizing research with respect to the specific statistical methodology used, a review of issues related to bias in the cognitive instrument under study, and a review of the demographic issues which have received the greatest emphasis in research on racial/ethnic subgroups, especially African American children.

The use of factor analysis as the methodology in this study is the most important consideration in evaluating the problem of equivalence of constructs. The

literature generated with respect to the use of factor analysis with the Wechsler Intelligence Scale for children during the past 50 years was reviewed comprehensively, therefore, initially, in this chapter. The two-factor model, which reflects Wechsler's organization of the test, was robustly supported over all three versions of the WISC for children across samples including those varying in racial/ethnicity status and disability status. The three-factor model, and four-factor models, by contrast, received less consistent support in racial/ethnically diverse samples and sample with disabilities.

A summary of issues related to bias in the use of the Wechsler Intelligence Scales for Children in the cognitive assessment of American-born, English-speaking subgroups followed using the seven categories outlined by Reynolds et al. (1999). It offers powerful empirical evidence from psychometric and statistical analyses to support the conclusion that the most prominent tests of cognitive abilities, including the WISC-III, provide equivalent validity across culturally diverse groups (Brown et al., 1999).

The final sections reviewed research related to the four demographic characteristics generating the largest number of studies involving African American children with disabilities. Although these characteristics are not directly relevant to the question of this study involving equivalence of constructs, they broaden and deepen understanding of issues related to the assessment of ethnic minority children.

CHAPTER III

METHOD

Participants

The WISC-III Standardization Sample

Archival data from the WISC-III standardization sample were obtained from The Psychological Corporation. Please refer to “Composition of the WISC-III” in Chapter I and Appendix A entitled “Characteristics of the Standardization Sample of the WISC-III” for details concerning characteristics and specifications for the entire WISC-III standardization sample. Permission was obtained to use subtest scores from the Caucasian sample of 1543 students without disabilities (CN) and the African American sample of 338 students without disabilities (AAN). Additionally, permission was granted to include subtest scores from a Caucasian mixed clinical validity sample of 207 students with disabilities (CD). This sample was drawn as part of the overall standardization sample data gathering process, but was not included in the actual norming of the instrument. The mixed CD sample included 77 students classified as having attention deficit disorders, 61 students classified as having learning disabilities, 26 students classified as having mental retardation, and 43 students classified as having speech/language disabilities.

The African American Sample of Children with Learning Disabilities

Archival data for the study were obtained from psychoeducational evaluations conducted in five school districts in two geographical regions, the Northeast and the South, during the 1991-1996 school years. WISC-III subtest scores were obtained for 1,058 African American children, ages 6-0 to 16-11, who were designated as having learning disabilities (AALD). The sample was drawn from urban and suburban communities. All of these children had been administered the WISC-III as part of an initial, triennial, or other psychoeducational evaluation. Information about each child, including date of birth, gender, ethnicity, and special education classification was obtained from school records. Traditional measures of socioeconomic status (i.e., parent education level or occupational status could not be accessed). It was estimated, however, that about 40% of the students in participating districts were eligible for free or reduced lunches, indicating this sample contained greater numbers of children coming from low socioeconomic status families than the WISC-III standardization sample.

Grade placement data presented two significant methodological problems. Some students in this sample were reported to be in “self-contained” (i.e., ungraded) special education classes. Information about the amount of time these students had spent in self-contained, versus less restrictive programs, and the precise definitions of how restrictive the self-contained programs were with respect to mainstreaming options, was not known for each student. Regarding high school students, status with respect to credits earned toward graduation was unknown. Even among those placed primarily in regular education programs, the school records sometimes indicated one

or more retentions had occurred. For all of the above reasons, it was decided grade placement information accessed could not be reported in this study.

Accurate information could not be obtained about how long each child had been placed in a special education program for several reasons. When a child transferred into a district included in the study, records from the previous district often were incomplete and did not specify consistently when the child had been classified initially. Even within the districts included in the study, information pertaining to date of initial classification could not be located consistently. In some cases, it was noted a child was found ineligible, declassified, and later reclassified, with incomplete reports available about the exact time periods covered.

Permission to access and utilize WISC-III subtest scores from school records of all students was obtained from central office administration in every participating school district. Because no identifying information was obtained regarding names of individual children, and because names of school systems, cities and states from which they were recruited will not be specified in the study, it was deemed unnecessary to obtain individual consent. School district administrators provided assurance that only properly credentialed and supervised school psychology staff members in participating districts had conducted assessments and that these staff had received adequate supervision regarding administration and scoring of protocols.

Criteria for inclusion in the study were threefold. All students must have received a WISC-III as part of their most recent initial or triennial or other psychoeducational evaluation. Additionally, they must have been classified as learning disabled. Also they must have been identified by the school district as

having African American ethnic status. Regulations developed by the state education departments in the two states from which this sample was drawn, which are in compliance with federal guidelines, were accepted as criteria for classification as having learning disabilities. In both cases, they include establishment of a severe discrepancy between ability and achievement by a multidisciplinary team.

All evaluations included individually administered tests of educational achievement. Credentialed school personnel conducted these tests. Achievement test scores were reviewed and recorded for the majority of the students, but not all of the children included in this sample. Selected subtests from the Woodcock-Johnson Psychoeducational Battery (WJ-R; Woodcock & Johnson, 1989) were reported most frequently, followed by selected Wechsler Individual Achievement Test subtest scores (WIAT; Wechsler, 1992), the Wide Range Achievement Test-Revised (WRAT-R; Jastak & Wilkinson, 1984), and selected subtests from the Kaufman Test of Educational Achievement (K-TEA; Kaufman & Kaufman, 1985).

Data sets collected for all students were carefully reviewed and incomplete data sets were eliminated where core subtests were missing either because a subtest did not yield a valid score or where no score was reported for a core subtest. List-wise, nearly 10% of the sample was excluded because of missing data. Nine hundred seventy-six children ($\underline{M} = 681$; $\underline{F} = 295$) completed the ten core subtests of the WISC-III and obtained scores reported as valid. Data from this group were used to test the two-factor model. The correlation matrix and standard deviations for this sample are reported in Appendix H. The Digit Span supplementary subtest and the ten core subtests were administered to 646 subjects ($\underline{M} = 453$; $\underline{F} = 193$) for whom scores were

reported as valid. Data from this group were used to test the three-factor model. Administration of the 12 subtests required to obtain the four index scores (which included the ten core subtests as well as Digit Span and Symbol Search) occurred in 171 cases ($\underline{M} = 124$; $\underline{F} = 47$) for whom scores were reported as valid. Data from this group were used to test the four-factor model.

Preliminary Data Treatment of All Samples

In preparation for all group comparisons involving the WISC-III Caucasian and African American standardization subsets, WISC-III clinical validity sample, and African American sample with learning disabilities it was first necessary to test whether the children from each age group from within each sample in each factor group could be collapsed into a single cross-age sample for that factor group. This was accomplished by investigating the feasibility of collapsing all of the scores across all 11 age groups for each factor sample in order to obtain a single covariance matrix for that factor model. The decision to create a covariance matrix for each group rather than a correlation matrix was guided by Byrne's (1994b) observation that "cross-group comparisons are generally only valid when covariance matrices are generated and analyzed" (p. 166). In order to accomplish this, it had to be determined whether the factor structure was invariant across all 11 age groups in each of the four samples used to test the two-factor model, three-factor model and four-factor model.

The method used to test for multigroup comparisons across the 11 age groups in each sample for each factor model involved maximum-likelihood partially restricted factor analysis using the EQS 5 for Windows (1995) program. As every age group was compared to another age group and subsequently collapsed into a cross

age group, there were two unique features in the age comparisons. First, the factor coefficients were constrained to be equal in both age groups being compared. Secondly, the parallel factor coefficients were estimated by the program simultaneously in a multigroup (i.e., two age levels run concomitantly) comparison. The constraints that were used to specify which parameters were held constant included specification that each core subtest was set to load on one and only one factor. The program, however, freely estimated the factor loadings. The factors were allowed to be correlated. The EQS program freely estimated the factor covariance.

For example, in the partially restricted factor analysis, the covariance matrix for the two-factor model was constructed initially for the sample of African American children with learning disabilities. The correlation matrix with standard deviations for the 11-year-olds was used to derive a covariance matrix for the 11-year-olds and the correlation matrix and standard deviations for the 12-year-olds was used to derive a covariance matrix for the 12-year-olds. Then a partially restricted maximum-likelihood factor analysis was conducted upon the covariance matrix of the scores for these two age groups simultaneously using EQS 5 for Windows (1995). Each core subtest was set to load on one and only one factor, but the factor loadings were not fixed and thus were freely estimated by the computer program. The factors were allowed to be correlated and the factor covariance freely estimated by the computer program as well. Each parallel factor coefficient was constrained to be equal to the other (e.g., Factor 1 to Verbal Subtest 1 for 11-year-olds must equal Factor 1 to Verbal Subtest 1 for the 12-year-olds). Goodness-of-fit indices were examined.

Provided the two groups were not discrepant, the data from the two groups were then collapsed and combined, creating a single covariance matrix from both the 11 and 12-year-olds. This covariance matrix was subsequently compared to the covariance matrix of the 10-year-olds in a partially restricted factor analysis with full constraints. Following this, the combined scores of the 10, 11, and 12-year-olds were collapsed and compared to the scores of the 13-year-olds. These procedures were repeated across the other seven age groups until a single cross-age covariance matrix was constructed for the entire sample of African American children with learning disabilities. This procedure was repeated involving the other three samples (i.e., the WISC-III Caucasian sample, the WISC-III African American sample, and the WISC-III mixed validity sample with disabilities) for the two-factor model. Subsequently, this same procedure was repeated for the four samples in constructing the covariance matrix for the three-factor and four-factor models as well. Where sample sizes were adequate within an individual age group in each sample of the factor models, the fit statistics were above .90, indicating the appropriateness of combining the entire sample as reported in Appendix E. The fit statistics for the combined cross-age groups for all of the factor models were above .90.

Instrument

The WISC-III is an individually administered assessment tool designed to measure intelligence of children ages 6-0 to 16-11 (Wechsler, 1991). The WISC-III consists of 10 core subtests and 3 supplementary subtests. A brief description of the WISC-III subtests is presented in Table 6.

Table 6

Descriptions of the WISC-III Subtests

Subtest	Description
Picture Completion	A set of colorful pictures of common objects and scenes, of which each picture is missing an important part, that the child identifies.
Information	A series of orally presented questions that tap the child's knowledge about common events, objects, places, and people.
Coding	A series of simple shapes (Coding A) or numbers (Coding B), each paired with a simple symbol. The child draws the symbol in its corresponding shape (Coding A) or under its corresponding number (Coding B), according to a key. Coding A and B are included on a single perforated sheet in the record form.
Similarities	A series of orally presented pairs of words for which the child explains the similarity of the common objects or concepts they represent.
Picture Arrangement	A set of colorful pictures, presented in mixed-up order, which the child rearranges into a logical story sequence.
Arithmetic	A series of arithmetic problems which the child solves mentally and responds to orally.
Block Design	A set of modeled or printed two-dimensional geometric patterns which the child replicates using two-color cubes.

Table 6 (continued)

Subtest	Description
Vocabulary	A series of orally presented words which the child orally defines.
Object Assembly	A set of puzzles of common objects, each presented in a standardized configuration, which the child assembles to form a meaningful whole.
Comprehension	A series of orally presented questions that require the child's Solving everyday problems or understanding of social rules and Concepts.
Symbol	A series of paired groups of symbols, each pair consisting of a Target group and a search group. The child scans the two groups And in indicates whether or not a target symbol appears in the Search group. Both levels of the subtest are included in a single response booklet.
Digit Span	A series of orally presented number sequences that the child Repeats verbatim for Digits Forward and in reverse order for Digits Backwards.
Mazes	A set of increasingly difficult mazes, printed in a response booklet, which the child solves with a pencil.

Note. From Manual for the Wechsler Intelligence Scale for Children- Third Edition

(p. 6), by D. Wechsler, 1991. San Antonio: The Psychological Corporation.

Copyright 1991 by the Psychological Corporation. Reprinted with permission.

The distribution of raw scores at each age level for each of the 13 subtests in the WISC-III battery was converted to a scaled score with a mean of 10 and a standard deviation of 3. Full Scale, Verbal, and Performance IQ's are obtained by summing subtest scores and referring to appropriate norm tables in Appendix A of the WISC-III manual. The Verbal, Performance, and Full Scale IQ's were developed to yield a mean of 100 and a standard deviation of 15. Additionally, four index scores, Verbal Comprehension, Perceptual Organization, Freedom for Distractibility, and Processing Speed are obtained utilizing scores from appropriate subtests and referring to the appropriate norm tables in Appendix A of the WISC-III manual. The index scores, congruent with the Verbal, Performance, and Full Scale IQ's also were constructed to have means of 100 and standard deviations of 15. Issues that are related to interpretation of index scores are reviewed in the section in Chapter I entitled "Use of Index Scores in Test Interpretation."

Although the essential structure of the test was retained in the WISC-III revision, some noteworthy changes were made, the most significant being the addition of a single new optional subtest, Symbol Search. In Appendix A the section entitled "Content Revision of the WISC-III" contains detailed information concerning Symbol Search. Symbol Search is relevant to this study because it combines with Coding to form the Processing Speed factor in the four-factor model.

Mazes, an additional supplementary subtest, was administered to less than 1% of the sample of African American children with learning disabilities included in this study. It was excluded from analysis in this study for two reasons. Practitioners in conducting psychoeducational assessments rarely use it (Glutting, Konold,

McDermott, Kush, & Watkins, 1995). Additionally, Mazes is not included in the calculation of Verbal, Performance, and Full Scale Intelligence Quotients and Index Scores (Konold, Kush, & Canivez, 1997; Wechsler, 1991).

The two most important goals in the development of the WISC-III were updating of norms and content revision. The section in Chapter I entitled “Wechsler Intelligence Scale for Children-III” and Appendix A entitled “The WISC-III” contains detailed information concerning renorming. Appendix A also contains detailed information regarding content revision.

Reliability

Extensive studies were reported in the WISC-III manual (Wechsler, 1991) regarding the short-term reliability of the instrument. The average internal consistency reliability coefficients for the FSIQs, VIQs, and PIQs generated from the 10 core subtests at 11 age levels were reported as .96, .95, and .91, respectively. Kamphaus (1993) reports the average subtest coefficients range from .70 - .87. Test-retest reliability was determined by comparing WISC-III scores of 353 subjects (drawn from 6 age groups) of the standardization sample who were retested in intervals of 12 to 63 days. The stability coefficient of the FSIQ was reported as .94 (Wechsler, 1991). The standard error of measurement of the FSIQ across all age levels was 3.20 (Wechsler, 1991).

Long-term stability of the WISC-III has been investigated in 3 recent studies, all involving children with disabilities. In reviewing the results of these studies, however, reliability is substantially affected by restriction in range. The IQs of children with learning disabilities frequently fall below average. Stavrou and

Flanagan (1996) reported Pearson product-moment correlation coefficients for FSIQ, VIQ, and PIQ of .82, .76, and .71, respectively, in a sample of 50 children with learning disabilities. The children were retested at a 3-year interval. The ethnicity of the sample was not specified. Zhu, Woodell, and Kreiman (1997) reported Pearson product-moment correlation coefficients for FSIQ, VIQ, and PIQ of .78, .79, and .70, respectively, in a sample of 60 children with learning disabilities. The retest interval range in this study was 32 - 48 months. Zhu et al., (1997) reported significant decreases in FSIQ, VIQ, and PIQ from first to second testing. The ethnicity of the sample was not specified.

Canivez and Watkins (1998) investigated long-term stability of the WISC-III for a heterogeneous sample of children with disabilities ($n = 667$). About 55% of the sample were children with learning disabilities. The sample included 98 African American children. The mean test-retest interval was 2.83 years (range .5 - 6.2 years). The Pearson product-moment correlation coefficients for the FSIQ, VIQ, and PIQ were .91, .87, and .87, respectively. Index score correlations were reported as well with significant correlations of .85, .87, .75, and .62, respectively. Test-retest reliability coefficients for subtests ranged from .55 for Symbol Search to .78 for Block Design.

Extensive studies of short-term and long-term stability thus have not been conducted to date with the WISC-III for different ethnic groups. Preliminary data available, however, suggest acceptable reliability, both short-term and long-term, and congruence with reliabilities reported in WISC-R studies. The reliability of the WISC-R across different ethnic groups was found to be consistent with the reliability

estimates of the WISC-R standardization sample (Suzuki, 1992). There was some evidence that long-term IQ score stability was greater for Caucasians than African Americans or Hispanics with disabilities (Elliott, S. N., Piersel, Argulewicz, Gutkin, & Galvin, 1985).

Validity

Concurrent Validity

Studies of the concurrent validity of the WISC-III Full Scale IQ's with other Wechsler Scales reveal correlations ranging from .85 - .88 (Wechsler, 1991). Correlations with other major intelligence scales reported in the WISC-III manual are only slightly lower. The correlation with the DAS General Cognitive Ability Score was .84 in the 8-9 age range and .91 in the 14-15 age range. The correlation with the SB-IV composite score was .83 in the 6-16 age range.

Construct Validity

The extensive investigation of the factor structure of the WISC-III standardization samples other national samples and clinical samples are summarized in detail in Chapter II in the section entitled "Factor Analysis of the WISC-III." Detailed information about clinical samples exclusively of African American children is reported in Chapter II in the section entitled "Summary of Studies Involving Children with Learning Disabilities." Detailed information about other samples of children with disabilities is reported in that section also.

Procedure

Overview of Structural Equation Models

Structural equation modeling (SEM), also known as covariance structure modeling (CSM), has achieved prominence as a research tool in the behavioral sciences during the past two decades (Kaplan, 1990; MacCallum, 1986; MacCallum, Rioznowski, & Necowitz, 1992; Tanaka, 1990). Terminology is fluid in this area of statistical research with respect to the use of SEM and CSM. In the present study the term SEM is used consistently.

“Structural equation modeling (SEM) is a statistical methodology that takes a hypothesis testing (i.e., confirmatory) approach to the multivariate analysis of a structural theory bearing on some phenomenon” (Byrne, 1994b, p. 3). The hypothesis involves “a specific pattern of relations among a set of measured variables (MVs) and latent variables (LVs)” (MacCallum, 1986, p. 107). “Drawing on knowledge of the theory, empirical research, or both ... [the researcher] postulates the linkage pattern a priori and then tests the hypothesis statistically” (Byrne, 1994b, p. 5). The fundamental unit of analysis in SEM, according to Tanaka (1990), is observed variances and covariances.

The structural theory often represents the processes that are causally generating the observations on multiple variables in the sample data (Bentler, 1988). In the context of structural equation modeling, the linkages between the factors and their measured variables are considered to represent a measurement model (Byrne, 1994b). Linkages exclusively among the latent variables comprise a structural model. The causal processes are represented by structural or regression equations which are

modeled in diagrams pictorially to enable the reader to understand the theory which they represent (Byrne, 1994b). A detailed description of EQS notations and path diagrams is located in Appendix C.

There are several advantages to SEM. First, “by demanding that the pattern of intervariable relations be specified a priori, SEM lends itself well to the analysis of data for inferential purposes” (Byrne, 1994b, p. 3). Secondly, it provides explicit estimates of parameters assessing measurement error (Byrne, 1994b). Additionally, it incorporates not only observed variables but also latent variables. Because of these features it has achieved a place of prominence in non-experimental research, wherein methods for testing theories have not been thoroughly developed (Bentler, 1980).

The present research was an example of nonexperimental research that was conducted to investigate latent variable structures in four samples across three models. The three models have generated considerable empirical research, although none is solidly grounded in a major theory. The present study undertook hypothesis testing of the structure of intervariable relationships of the WISC-III subtests across samples differing in ethnicity and disability status. The basic question was whether the factor structure was invariant across the three well-researched models, each specifying a different set of latent variables.

The methodology of SEM appears well suited to address this question. One of the most typical applications of SEM is an application involving model fitting to sample data (MacCallum, 1986). A researcher uses SEM “to specify a model that is consistent with the observed data” (Kaplan, 1990, p. 140). The researcher employs confirmatory factor analysis to test the model fit of the sample data. When the

researcher evaluates the solution obtained from the model fitting to the sample data in confirmatory factor analysis, parameter estimates and goodness of fit typically are investigated (MacCallum, 1986). The goodness of fit criteria “guide the search for a best fitting solution” (Loehlin, 1992, p. 60) and “they evaluate the solution when it is obtained” (Loehlin, 1992, p. 60) for the model as a whole. The goodness-of-fit statistics provide feedback about the degree to which the factor structure model fits or reproduces the covariances among the data.

With data generated in the behavioral sciences, a perfect fit between the hypothesized model and the observed data is highly unlikely (Byrne, 1994b). The discrepancy that emerges between the observed sample covariances and the model estimates of the population covariances is termed the residual (MacCallum, 1986). In situations where large residuals emerge, reflecting poor fit, a specification search (Leamer, 1978; Long, 1983) often is conducted to determine if modification of the model would improve its fit to the data (MacCallum, 1986). After the identification of specification errors reflecting the lack of correspondence “between a proposed model and the true model characterizing the population and variables under study” (MacCallum, 1986, p.108), correction procedures are imposed on the model. The goal of the respecification is to arrive at a model that more accurately represents the relationships among the MVs and LVs in the population (MacCallum, 1986).

Perhaps the most commonly utilized statistical criterion used to evaluate model fit to data has been the chi-square likelihood ratio statistic. The chi-square likelihood ratio statistic measures “the closeness of fit between the sample covariance matrix and the fitted covariance matrix, serving therefore as an indicator of overall

model fit” (Byrne, 1994a, p. 293). From this perspective Joreskog and Sorbom (1993) characterize the chi-square value as actually being a measure of badness-of-fit. This is because the smaller the value of the chi-square statistic, the better is the fit. Conversely, a large value of the chi-square statistic is associated with the need to modify the model in order to attempt to better fit the data (Joreskog & Sorbom, 1993).

The relative fit of theoretically competing models can be evaluated with the likelihood ratio chi-square test. “This method is most useful when one model is hierarchically nested within the other [i.e., when one contains the same set of parameters as the other, plus some additional parameters]” (Breckler, 1990, p. 262). “Large drops in the statistic indicate that the changes made in the model represent a real improvement” (Joreskog & Sorbom, 1993, p. 29). The chi-square statistic has been found to be very sensitive to sample size, however. It typically is significant for most researchers’ models, suggesting the models should be rejected, even when an inspection of their residuals would reveal trivial differences (Mulaik et al., 1989).

For that reason the chi-square statistic has been augmented by numerous alternative fit indices developed during the past 20 years (Marsh, Balla, & McDonald, 1988). Appendix D contains a detailed description of the fit statistics reported using the EQS Program. The formulae for all fit indices reported in this study are presented in Table 7.

The fit indices were designed “to avoid some of the problems of sample size and distributional misspecification on evaluation of a model” (Hu & Bentler, 1998, p. 425). Hu and Bentler characterized them as summary statistics. The fit indices

“generally quantify the extent to which the variation and covariation in the data are accounted for by a model” (Hu & Bentler, p. 426). Fit indices often are categorized as being stand-alone or absolute and incremental (Hu & Bentler, 1998; Marsh et al., 1988). The possible values for most of these indices range between zero and unity with zero reflecting a complete lack of fit and unity reflecting perfect fit (Mulaik et al., 1989). However, some indices have a wider range and can have negative values and positive values greater than 1. When maximum-likelihood (ML) criteria are employed the fit indices have a distinct statistical advantage. ML “yields a quantity that is approximately distributed as chi-square, permitting statistical tests of goodness of fit” (Loehlin, 1992, p. 60). This advantage is important in situations where the model fit is stressed.

Hu and Bentler (1998), having investigated the sensitivity of ML fit indices to model misspecification under conditions involving variations in sample size and distribution, recommend the use of a “two-index presentation strategy for researchers” (p. 447). Additionally they recommend that parameter estimates be examined relating to substantive issues and that residual covariances be examined. The TLI and CFI were among the most highly recommended indices with ML when paired with the Standardized Root-Mean-Square Residual (SRMR). Hu and Bentler reported, however, “the average absolute standardized residual computed by EQS ... has an identical rationale and should perform the same as SRMR” (p. 447).

In summary, the TLI and CFI offer somewhat different advantages. The TLI incorporates a penalty function for nonparsimonious models but is less preferable for sample sizes less than 250 (Hu & Bentler, 1998). The CFI does not incorporate a

penalty function for nonparsimonious models but is more preferable for sample sizes under 250 (Hu & Bentler). In the current study several samples are under 250. Additionally, three models are under study and parsimony may be an issue regarding goodness-of-fit. Finally, the TLI was reported in CFA analyses of the entire WISC-III standardization sample. The TLI, therefore, serves as an important index in comparing results across all of the samples from the present study and the entire standardization sample. For these reasons both fit indices will be reported. In the current study the baseline model used in computing the TLI is a null model with no latent factors. In keeping with Hu and Bentler's recommendations, the average absolute standardized residuals will be examined and included as well.

Byrne's (1994b) criterion of $r = .90$ and above will be used with respect to the TLI and CFI. Although this criterion is somewhat more liberal than the $r = .95$ or above criterion suggested by Hu and Bentler, it is considered appropriate in the current research because there are considerable differences in sample sizes. These are known to have powerful effects on both model error and sampling error (Hu & Bentler, 1998; MacCallum et al., 1992), as well as the interaction between model error and sampling error (Silvia, 1988). Hu and Bentler suggested "a cutoff value close to .08 for SRMR" (Hu & Bentler, 1998, p. 449). In the current study, the examination of average absolute standardized residuals computed by EQS reflected Hu and Bentler's caution about the upper bound of SRMR values that should be accepted.

Table 7

Goodness of Fit Indices for Maximum Likelihood Estimation Procedures

LISREL GFI Fit Index:	$1 - [\text{tr}(E^{-1} * S - I)^2 / \text{tr}(E^{-1} * S)]^2$
	E = the fitted matrix, S = the covariance matrix of the observed variables, I = identity matrix
LISREL AGFI Fit Index:	$1 - [[k(k + 1) / 2(\underline{df}_m)] * (1 - GFI)]$
	k = the number of observed variables
McDonald Fit Index (MFI):	$\text{Exp} [-.5 * (\chi^2_m - \underline{df}_m) / (\underline{N}-1)]$
Root Mean Squared	$[(\Sigma s) / k(k + 1)]^{.5}$
Residual (RMSR):	s = the residual covariances
Bentler-Bonett Normed Fit	$(\chi^2_b - \chi^2_m) / \chi^2_b$
Index (NFI):	
Tucker-Lewis Index (TLI),	$[(\chi^2_b / \underline{df}_b) - (\chi^2_m / \underline{df}_m)] / [(\chi^2_b / \underline{df}_b) - 1]$
also called the Bentler-	b = independence model χ^2 , m = fitted model χ^2
Bonett Nonnormed Index:	
Bollen Fit Index (IFI):	$(\chi^2_b - \chi^2_m) / (\chi^2_b - \underline{df}_m)$
comparative fit index (CFI):	$1 - [(\chi^2_m / \underline{df}_m) / (\chi^2_b / \underline{df}_b)]$
	If $(\chi^2_m / \underline{df}_m) < 0$, then $(\chi^2_m / \underline{df}_m)$ is set to equal 0. If $(\chi^2_b / \underline{df}_b) < 0$, then $(\chi^2_b / \underline{df}_b)$ is set to equal 0

Note. The goodness-of-fit indices included in this study were generated by the EQS 5

for Windows (1995) program.

In the EQS program, assessment of parameter misspecifications is accomplished by means of the Lagrange Multiplier (LM) Test. Results of the LM Test are presented as univariate test statistics and multivariate test statistics. Byrne (1994b) recommended use of the multivariate test statistics over those of the univariate test statistics because the latter test restrictions in the model independently. Correlations among particular variables are not taken into account. The multivariate test statistics, which take the correlations into account, typically produce fewer suggestions of malfitting parameters.

Byrne (1994b) cautions that a researcher cannot rely solely on output from the LM Test to make modifications in a model. This is because the LM Test, which is based on statistical criteria alone, evaluates constraints in virtually any fixed parameter without attending to the theoretical rationale underlying the model that is respecified. "Model respecification in which certain parameters have been set free must be substantiated by sound theoretical rationale; it also demands that attention be paid to the issue of identification" (Byrne, 1994b, p. 48).

In the current study, the LM Test was used to conduct specification searches in each case where the three models are imposed on sample data in fully restricted factor analyses. Because previous research conducted regarding observed variables and latent variables suggested the variables are intercorrelated, in the present study the multivariate test statistic was relied upon in specification searches. Model respecification involving freeing certain parameters was carefully considered only within the rationales observed from empirical research on the two-factor, three-factor, and four-factor models. The results of the LM Tests are discussed in Chapter IV.

Design of the Current Study

The generalizability of the factor structure of the Caucasian nondisabled subset of the WISC-III standardization sample to an African American learning disabled sample was investigated utilizing CVCSM, in an extension of the work of Allen and Thorndike (1995a, 1995b). This study attempted to address a methodological issue raised by Allen and Thorndike (1995a, 1995b) in relation to the factor analysis of the WISC-III standardization sample reported in the WISC-III manual (Wechsler, 1991), namely that “no direct comparisons were conducted between groups to examine factor invariance” (p. 8). The purpose of this investigation was to determine the degree to which the factor structure of the WISC-III could be shown to be invariant across ethnicity and disability status using the three models which have generated the greatest amount of research to date. The three models were included of necessity, because there is no consensus about which of the three models fits best across diverse samples.

In this investigation a partially restricted factor analysis of the Caucasian subset of children without disabilities (CN) served as a model in a series of fully restricted factor analyses, forcing the solution to contain two, then three, and finally four factors. These involved the WISC-III sample of African American children without disabilities (AAN), the WISC-III Caucasian clinical validity sample of children with disabilities (CD), and the sample of African American children with learning disabilities (AALD).

Initially, a partially restricted maximum-likelihood (ML) factor analysis of the 10 core subtest scores of the Caucasian unrestricted subset of the WISC-III

standardization sample was conducted using EQS 5 for Windows (1995). There were two reasons why ML was selected. First, it was the method used most frequently in the factor analyses of the WISC-III standardization sample cited in the WISC-III manual (Allen & Thorndike, 1995a, 1995b; Kush et al., 1997; Kush et al., in press). Second, it was directly congruent with the fitting function utilized in the fully restricted procedures. The use of ML therefore eliminated a source of methodological inconsistency both within the study and in comparing results to other relevant studies. Additionally, Hu and Bentler (1998) reported that most fit indices obtained from ML are less likely to be influenced by irrelevant effects and to depart from true-population values than from other methods they studied.

In each investigation of a factor model, simple structure was imposed wherein each subtest defining a factor in the model was forced to be associated with that factor alone. The structure for the two-factor model, for example, was set as defined by Wechsler's V/P dichotomy, forcing the solution to contain two factors in a simple factor structure. For purposes of this study, a partially restricted factor analysis meant that each core subtest was set to load on one and only one factor, but the factor loadings were not restricted and thus were freely estimated by the computer program. Additionally, factors were allowed to be correlated. The partially restricted factor solution obtained from this factor analysis of the CN subset subsequently served as the model in the series of fully restricted analyses. For purposes of this study, a fully restricted factor analysis meant that each core subtest was required to load on one and only one factor, and further that the factor loadings were fixed to prespecified values and were not estimated by the computer program. The intercorrelations among the

two factors also were fixed. All of the other parameters, such as factor variances and error variances, were left free to be estimated by the computer program. All of the restricted factor analyses were conducted utilizing the structural equation modeling program EQS 5 for Windows (Bentler, 1995).

The factor structure and factor loadings that were obtained from the partially restricted factor analysis of the CN subset were imposed as fixed parameters upon the covariance matrix from the AAN subset of children. This analysis tested CN-AAN stability. It was expected to reflect differences due to racial effects. Then the factor structure and factor loadings that were obtained from the partially restricted factor analysis of the Caucasian unrestricted subset were imposed as fixed parameters upon the covariance matrix from the CD clinical validity sample in a fully restricted factor analytic model. This tested CN-CD stability, which was expected to reveal the structural effects of disability status. Further, the factor structure and factor loadings that were obtained from the partially restricted factor analysis of the CN subset were imposed as fixed parameters upon the covariance matrix from the sample of AALD children with learning disabilities in the fully restricted factor analytic model. This tested the combined effects of race and disability status. The Lagrange Multiplier (LM) Test was utilized additionally in the examination of the results of these fully restricted factor analyses. The purpose of the LM Test was to determine whether freeing certain parameters or adding paths would lead to a better fitting model (Byrne, 1994b).

Furthermore, partially restricted factor analysis of the scores of the AAN subset of the WISC-III standardization sample for the 10 core subtests was then

conducted, forcing the solution to contain two factors. The factor structure and factor loadings that were obtained from this partially restricted factor analysis of the AAN subset were imposed as fixed parameters upon the covariance matrix from the sample of AALD children in a fully restricted factor analytic model. By using a model developed on AANs, this comparison tested the effect of learning disability status within the African American ethnic group. Goodness-of-fit statistics were generated. The LM Test was conducted.

A partially restricted factor analysis of the scores of the CD subset of the WISC-III validation sample for the 10 core subtests was then conducted, forcing the solution to contain two factors. The factor structure and factor loadings that were obtained from this partially restricted factor analysis of the CD subset were imposed as fixed parameters upon the covariance matrix from the proposed sample of AALD children in a fully restricted factor analytic model. This tested the effects of ethnicity within the groups of children with disabilities. Goodness-of-fit statistics were generated. The LM Test was conducted.

Identical procedures were repeated subsequently, forcing a three-factor solution and then a four-factor solution. Figure 1 illustrates the entire sequence of analyses. Note that the capital letters in Figure 1 correspond to the stages of cross-validation summarized in Table 8.

For all fully restricted factor analyses, all goodness-of-fit indices were reported. The likelihood chi-square ratio was reported. The following additional statistics were reported as well: the LISREL Goodness-of-Fit (GFI); the LISREL Adjusted-Goodness-of-Fit index (AGFI); the McDonald Fit index (MFI); the LISREL

Root Mean Squared Residual (RMSR); Bentler-Bonett Normed Fit index (NFI); the Tucker Lewis index (TLI) which is frequently reported as the Bentler-Bonett Nonnormed Fit index (NNFI); the Bollen Incremental Fit index (IFI); and the comparative fit index (CFI). The inclusion of all indices was necessary because there is no consensus about which index or indices provide a superior fit statistic in cross-validation designs (Allen & Thorndike, 1995a, 1995b; Bentler, 1989; Loehlin, 1992).

The fit statistics from the various comparisons of models were examined, with particular focus on the TLI and CFI. An overview of these measures was used to investigate the generalizability of the factor structure across different groups. Of particular interest is the comparison of the average change in fit statistics across ethnicity and disability status. This was examined initially by comparing the average change in fit statistics for the WISC-III AAN sample and the AALD sample. Following this comparison, the average change in fit statistics for the CD clinical validation sample with the AALD sample was reviewed. Subsequently, the differences were studied to determine whether ethnicity or disability status cause more stress in the factor structure across each model as reflected in deterioration in value of fit statistics. Reduction in the chi square and RMSR, and increases in the TLI and CFI indicate better model fit (i.e., less stress). Increases in chi square and RMSR and decreases in TLI and CFI are associated with less adequate model fit (i.e., greater stress). Finally, the model evaluation statistics were reviewed for the four samples across the three factor models in order to determine which model provides the best fit for each sample.

Figure 1

The Sequence of Factor Analyses Across Ethnicity and Disability Status

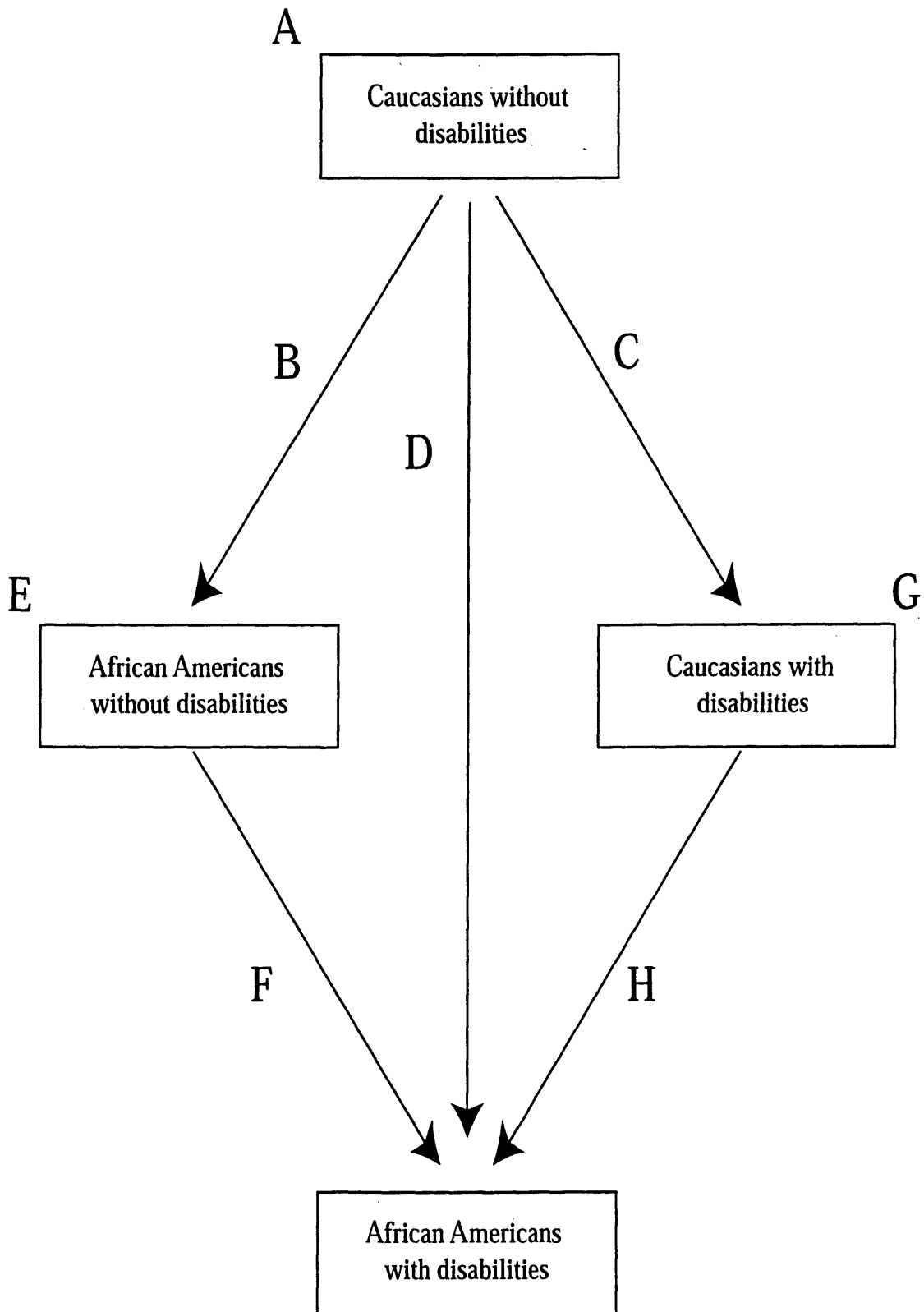


Table 8

Methodology: Stages in Cross-Validation Across Ethnicity and Disability Status for All Samples

Step	Procedure
A.	Determine the structure of the two-factor model for the WISC-III Caucasian nondisabled subset
	1. Partially restricted factor analysis of the WISC-III Caucasian nondisabled subset
	2. Examine fit statistics
B.	Cross validate to African American nondisabled subset
	1. CVCSM of Caucasian nondisabled subset structure and factor loadings imposed upon African American nondisabled subset in fully restricted factor analysis
	2. Examine fit statistics for appraisal of discrepancies
C.	Cross validate to Caucasian disabled subset
	1. CVCSM of Caucasian nondisabled subset structure and factor loadings imposed upon Caucasian disabled subset in fully restricted factor analysis
	2. Examine fit statistics for appraisal of discrepancies
D.	Cross validate to African American Learning Disabled Children
	1. CVCSM of Caucasian nondisabled subset structure and factor loadings imposed upon the proposed sample of African American learning disabled children in fully restricted factor analysis
	2. Examine fit statistics for appraisal of discrepancies

Table 8 (continued)

Step	Procedure
E.	Determine structure of the WISC-III African American nondisabled subset
	1. Partially restricted factor analysis of WISC-III African American nondisabled subset
	2. Examine fit statistics
F.	Cross validate to African American disabled subset
	1. CVCSM of African American nondisabled structure and factor loadings imposed upon proposed sample of African American learning disabled children in fully restricted factor analysis
	2. Examine fit statistics for appraisal of discrepancies
G.	Determine structure of the WISC-III Caucasian disabled subset
	1. Partially restricted factor analysis of WISC-III Caucasian disabled subset
	2. Examine fit statistics
H.	Cross validate to African American disabled subset
	1. CVCSM of Caucasian disabled structure and factor loadings imposed upon proposed sample of African American learning disabled children in fully restricted factor analysis
	2. Examine fit statistics for appraisal of discrepancies
I.	Repeat the above procedures for the three-factor and four-factor models
	1. Discuss overall pattern of fits and discrepancies in overview of models

Note. The capital letters denote steps followed in the cross-validation methodology and the Arabic numbers denote procedures outlined under each step.

CHAPTER IV

FINDINGS

Introduction and Organization

The summary of the descriptive statistics for samples are reported initially, followed by the results of the comprehensive data analysis for the two-factor model, the three-factor model, and lastly, the four-factor model. Discussion of the results for all three models follows these summaries. Subsequently, the fit of the models is examined for the four samples across models to determine which model fits the data best for each sample. Finally, factor loadings from other samples involving African American children with disabilities are compared to those of the AALD sample and where appropriate to the CD sample.

Data are reported within each factor model using the following format. The path diagram for the factor model and the results of the partially restricted factor analysis of the four samples is reported initially, including path and factor coefficients and targeted fit indices. The targeted fit statistics obtained when the results of the partially restricted factor analysis of the WISC-III Caucasian sample without disabilities were imposed as a fully restricted factor analysis on the WISC-III African American sample without disabilities (AAN), the WISC-III Caucasian subset with disabilities (CD), and the sample of African American students with learning disabilities (AALD) is reported in a table, including the chi square (χ^2), degrees of

freedom (df), standardized root-mean squared residual (SRMSR), average off-diagonal absolute standardized residual (Average R), Tucker-Lewis index (TLI), and comparative fit index (CFI). The eight fit indices outputted by EQS are presented for each partially restricted and fully restricted procedure in Appendix F. The discussion of results focuses upon the TLI, the CFI and the residuals, especially, for reasons explained in Chapter III.

Following these reports, the results of the application of the partially restricted factor analysis of the WISC-III AAN sample to the AALD sample in a fully restricted factor analysis are presented in terms of targeted fit statistics. The eight fit indices outputted by EQS are presented for each partially restricted and fully restricted procedure in Appendix F. The discussion focuses upon the TLI, CFI, and residuals.

Following these analyses, the results of the partially restricted factor analysis of the CD sample imposed on the AALD sample are reported in terms of targeted fit statistics. The eight fit indices outputted by EQS are presented for each partially restricted and fully restricted procedure in Appendix F. The discussion focuses upon the TLI, CFI, and residuals.

A summary of the analyses for the entire model follows. Areas of stress in the application of the model to these samples are reviewed. A summary of results across models concludes the chapter.

Constraints on Publication of Data

The agreement with the Psychological Corporation regarding WISC-III standardization sample data explicitly forbids the reporting of the means, standard deviations, correlation matrices and covariance matrices from the Caucasian WISC-

III sample without disabilities, the African American WISC-III sample without disabilities, and the WISC-III clinical validation sample of Caucasian children with disabilities. These data are considered proprietary and must be obtained directly from the Psychological Corporation.

The WISC-III manual reports the mean and standard deviation for the entire standardization sample calibrated to be 100, and 15, respectively. The standard deviations and correlation matrices for the CN sample, the AAN sample, and the CD sample were computed from subtest scores. It is important to note that the standard deviations for the PIQ and FSIQ for the CN and AAN samples were very similar to one another and did not differ greatly from the standard deviation calibrated for the entire standardization sample. There was slightly more variance noted in the standard deviation for the VIQ for the CN sample than the AAN sample. The standard deviations for the CD sample were considerably larger for VIQ, PIQ, and FSIQ than the standard deviations generated for the CN and AAN samples.

Descriptive Statistics

Data from WISC-III samples were used in this study. The WISC-III Caucasian standardization sample of 1,543 children without disabilities consisted of 771 boys and 772 girls. The WISC-III African American standardization sample of 338 children without disabilities consisted of 171 boys and 167 girls. Both samples were stratified. The WISC-III Caucasian validity sample of 207 children with disabilities consisted of 162 boys and 45 girls. It was not stratified.

The sample of 976 African American children with learning disabilities included in the factor analysis of the two-factor model consisted of 681 boys and 295

girls. For the three-factor model the sample of 646 children included 453 boys and 193 girls. The sample of 172 children used in the analysis for the four-factor model consisted of 125 boys and 47 girls. The age ranges of children in the entire African American sample with learning disabilities was 6 years, 0 months to 16 years, 11 months ($M = 9$ years, 4 months).

Descriptive statistics for the entire AALD sample for Verbal, Performance, and Full Scale IQ's of the WISC-III and individual subtests are reported in Table 9. The statistics reported reflect the averages obtained from the total number of valid scores reported for each subtest and composite IQ.

The standard deviations for the VIQ, PIQ, and FSIQ of the African American sample with disabilities were smaller than those reported for the CN and AAN samples. They were considerably smaller than those reported for the CD mixed validity sample. The standard deviations for the subtest scores for the CD sample reflected the widest variability among the four samples. The standard deviations of the majority of the subtests of the entire data set of the AALD sample were overall close to 3, consistent with the range reported for the WISC-III standardization sample. The most notable exceptions, reflecting greater restriction in range, were Arithmetic, Information, and Vocabulary.

The means for WISC-III VIQs, PIQs, FSIQs, index scores, and subtests for the four samples of African American children with disabilities were examined as reported in Table 10. Subtest means for the four clinical samples were compared. Consistently across all samples, well below average mean scores on Information and

Arithmetic were noted. These findings were consistent with WISC-R studies reviewed by Kaufman (1982) for samples of children with learning disabilities.

Table 9

Standard Score Means, Standard Deviations, and Ranges for WISC-III Verbal IQ, Performance IQ, Full Scale IQ, and Subtests for the Entire Sample of African American Children with Learning Disabilities (n = 1,058)

Subtest	Mean	SD	Range	
			Minimum	Maximum
Arithmetic	7.27	2.48	1	16
Block Design	7.63	3.08	1	19
Coding	8.29	3.19	1	19
Comprehension	7.89	2.96	1	17
Digit Span	7.95	2.87	1	19
Information	6.93	2.58	1	16
Object Assembly	7.75	3.07	1	19
Picture Arrangement	7.44	3.10	1	19
Picture Completion	8.08	2.93	1	17
Similarities	7.51	2.85	1	17
Symbol Search	7.89	3.25	1	17
Vocabulary	7.28	2.65	1	18
Full Scale IQ	84.81	11.11	53	133
Verbal IQ	85.42	11.78	52	131
Performance IQ	86.99	12.95	53	130

Note. Missing scores in individual profiles were found with respect to every variable.

The statistics presented in this table represent averages of the total number of scores obtained for each variable.

Table 10

Standard Score Means for WISC-II Verbal IQ, Performance IQ, Full Scale IQ, Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, Processing Speed and Subtests for Four Clinical Samples of African American Children

Subtest	<u>Means</u>			
	Sample 1	Sample 2	Sample 3	Sample 4
Verbal IQ	85.42	84.49	72.4	80.98
Performance IQ	86.99	85.02	76.2	82.08
Full Scale IQ	84.81	83.34	72.3	79.84
VC Index	83.55	85.81	70.8	83.16
PO Index	86.41	85.36	74.0	82.70
FD Index	83.91	85.34	82.3	---
PS Index	89.23	91.89	---	---
Information	6.93	7.12	5.1	6.49
Similarities	7.51	7.64	4.8	6.92
Vocabulary	7.28	6.99	4.3	9.71
Comprehension	7.89	7.54	4.1	7.45
Arithmetic	7.27	6.91	6.6	5.78
Digit Span	7.95	7.58	7.2	---
Picture Completion	8.08	7.95	5.1	7.77
Picture Arrangement	7.44	7.37	6.3	6.63
Block Design	7.63	6.92	4.9	6.35
Object Assembly	7.75	7.38	5.6	7.02
Coding	8.29	8.19	8.1	7.53
Symbol Search	7.89	8.28	---	---

Note. The empty cells reflect data not collected in two of the studies. Sample 1 = African American children with learning disabilities, Shindelman; Sample 2 = Kush et al.; Sample 3 = Slate & Jones; Sample 4 = Kush & Watkins.

Investigation of the Two-Factor Model

Overview

The ten core subtests of the WISC-III corresponding to the two-factor theoretical model are: Information, Similarities, Vocabulary, Comprehension, Arithmetic, Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding. The path diagram representing the theoretical structure of the two-factor model of the WISC-III is reported in Figure 2.

Partially Restricted Factor Analyses of the Caucasian Sample (CN), the African American Sample (AAN) of the WISC-III Standardization Sample, the Caucasian Validity Sample (CD), and the African American Sample of Children with Learning Disabilities (AALD) for the Two-Factor Model

Initially, a partially restricted maximum-likelihood factor analysis was conducted using the covariance matrices generated from the 10 core subtests (specified in the preceding section) for the WISC-III CN sample of 1,543 children, the African American sample (AAN) of 338 children, the Caucasian validity sample (CD) of 207 children with disabilities, and the African American sample of 976 children with learning disabilities (AALD). The structure coefficients for the subtests and the factor covariance for the four samples are reported in Table 11. Model evaluation statistics for the four samples are reported in Table 12 for the two-factor model. The path diagrams for the two-factor model for each sample including factor loadings, factor covariances, error variances for the subtests, and selected fit statistics are reported in Figures 3-6.

Figure 2

Hypothesized First-Order CFA Model for the Two-Factor Model

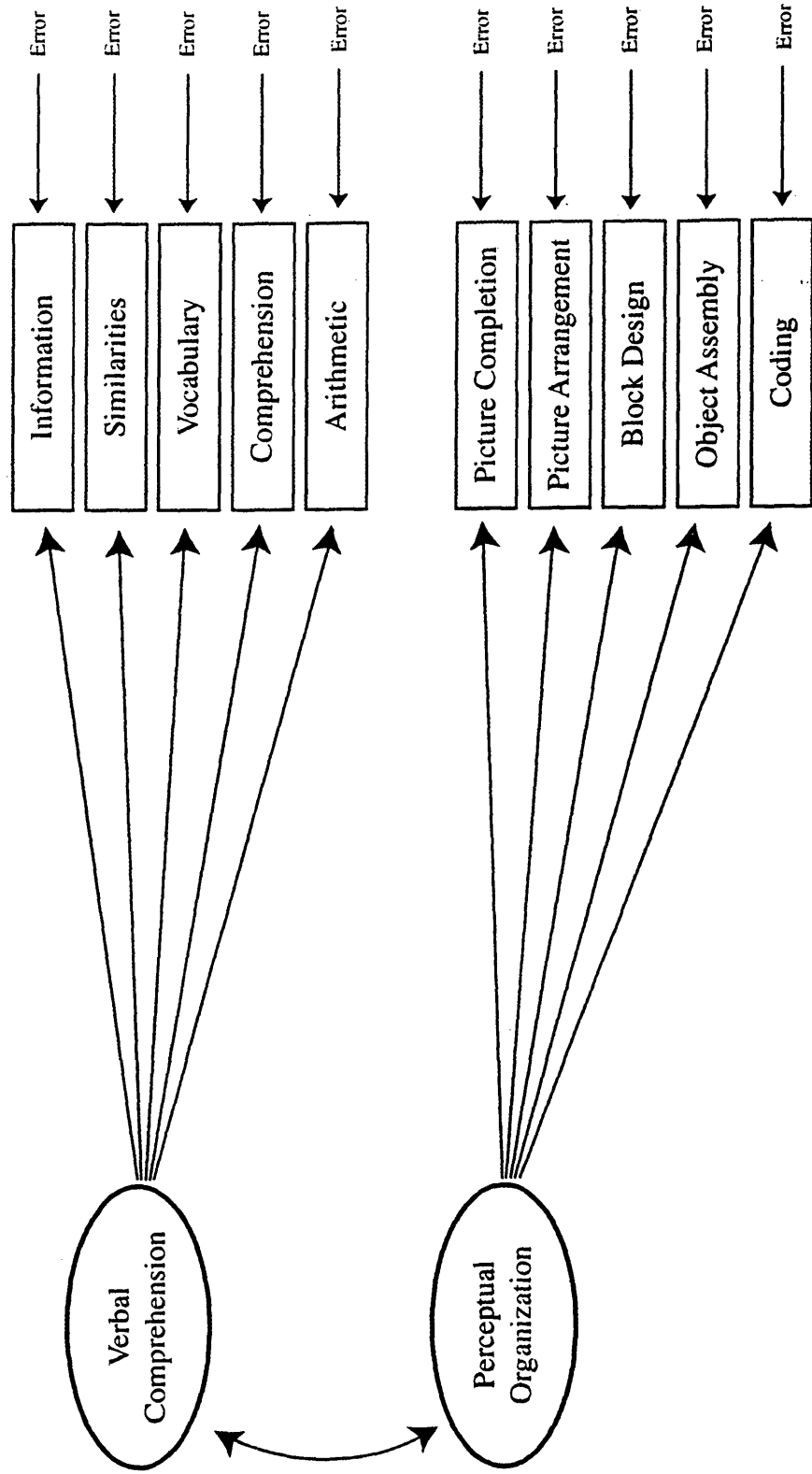


Table 11

Structure Coefficients Reported for All Samples in the Partially Restricted Factor Analysis for the Two-Factor Model

Subtests	Structure Coefficient			
	CN	AAN	CD	AALD
	(<u>n</u> = 1,543)	(<u>n</u> = 338)	(<u>n</u> = 207)	(<u>n</u> = 976)
	Factor 1 Verbal Comprehension			
Information	.79	.75	.81	.59
Similarities	.79	.77	.82	.70
Vocabulary	.81	.83	.88	.79
Comprehension	.69	.74	.83	.70
Arithmetic	.65	.64	.75	.57
	Factor 2 Perceptual Organization			
Picture Completion	.60	.69	.81	.62
Picture Arrangement	.50	.57	.76	.47
Block Design	.77	.79	.77	.59
Object Assembly	.70	.67	.79	.65
Coding	.36	.40	.55	.28
Factor Covariances F_1, F_2	.72	.76	.84	.51

Note. CN = WISC-III Caucasian Sample Without Disabilities; AAN = WISC-III

African American Sample Without Disabilities; CD = WISC- III Caucasian Validity

With Disabilities; AALD = African American Sample With Learning Disabilities.

Table 12

Model Evaluation Statistics for the Partially Restricted Factor Analysis for the Two-Factor Model

Sample	Goodness-of-fit statistics				
	χ^2	df	Average R	TLI	CFI
WISC-III Caucasian subset (<u>N</u> = 1,543)	252.73	34	.04	.95	.96
WISC-III African American subset (<u>N</u> = 338)	71.61	34	.03	.96	.97
WISC-III Caucasian subset with Disabilities (<u>N</u> = 207)	84.02	34	.04	.95	.96
African American Sample with Learning disabilities (<u>N</u> = 976)	165.08	34	.04	.92	.94

Note. Average R = Average off-diagonal absolute standardized residual;

TLI = Tucker-Lewis index; CFI = comparative fit index.

Figure 3

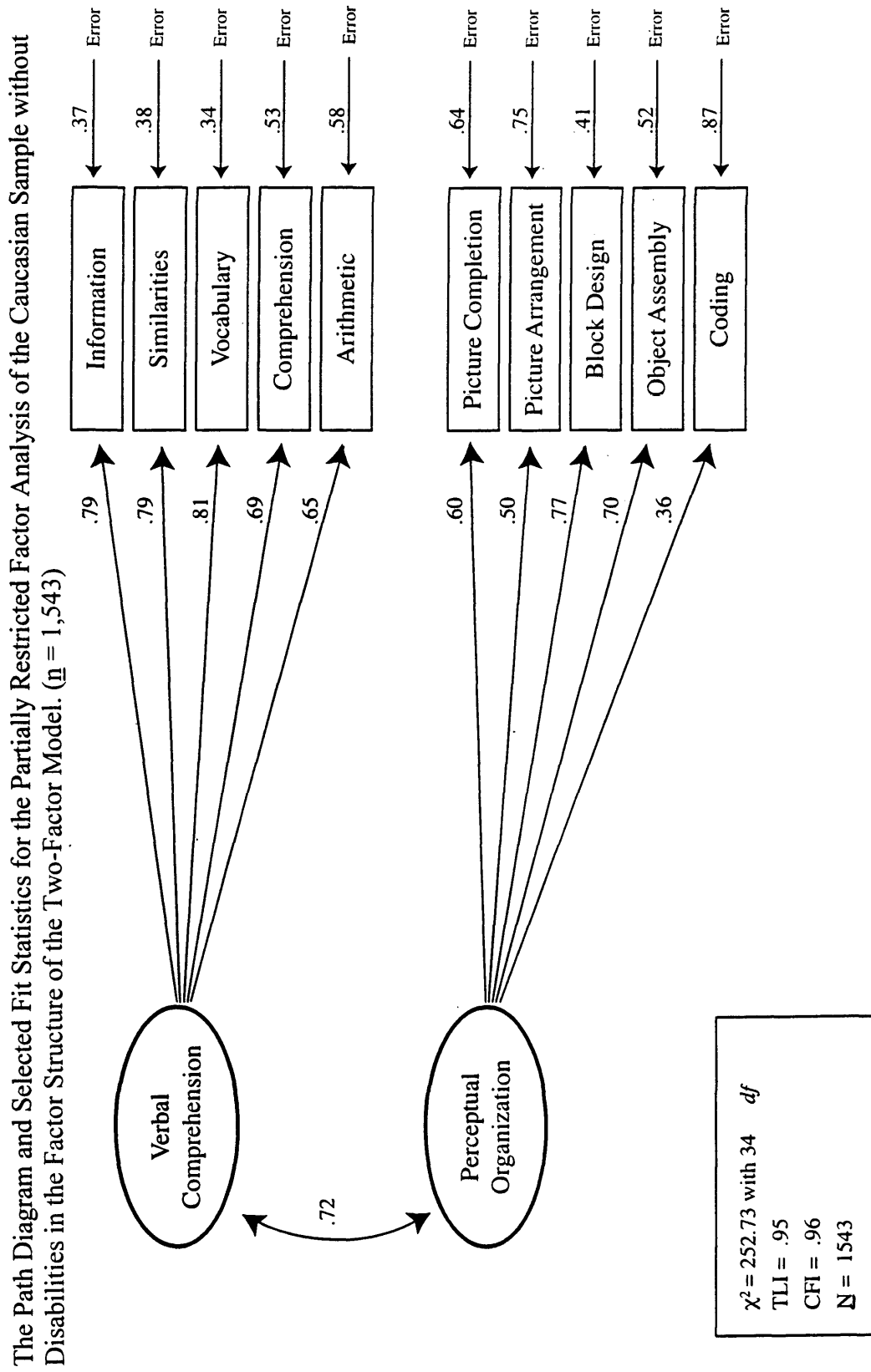


Figure 4

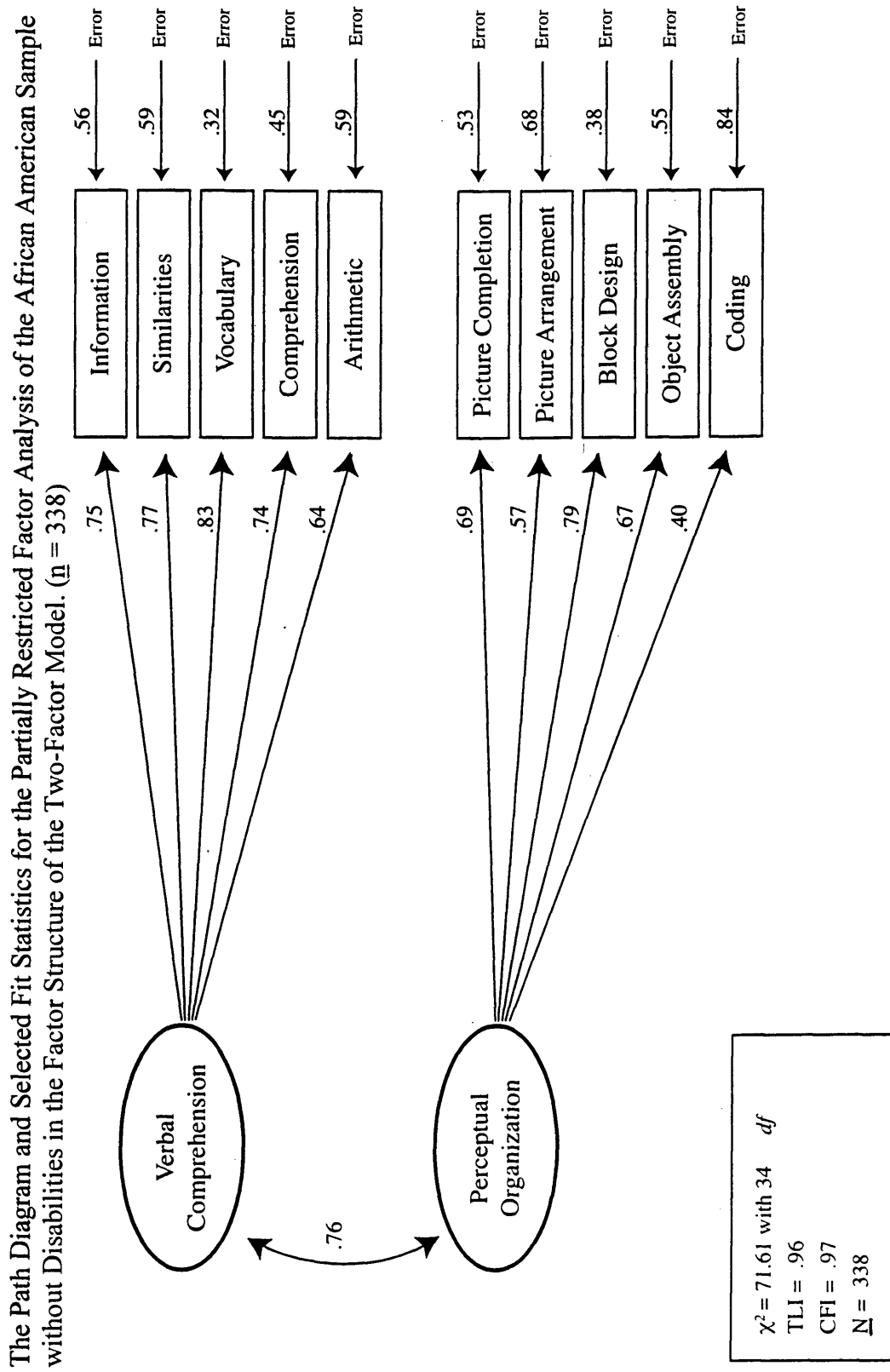


Figure 5

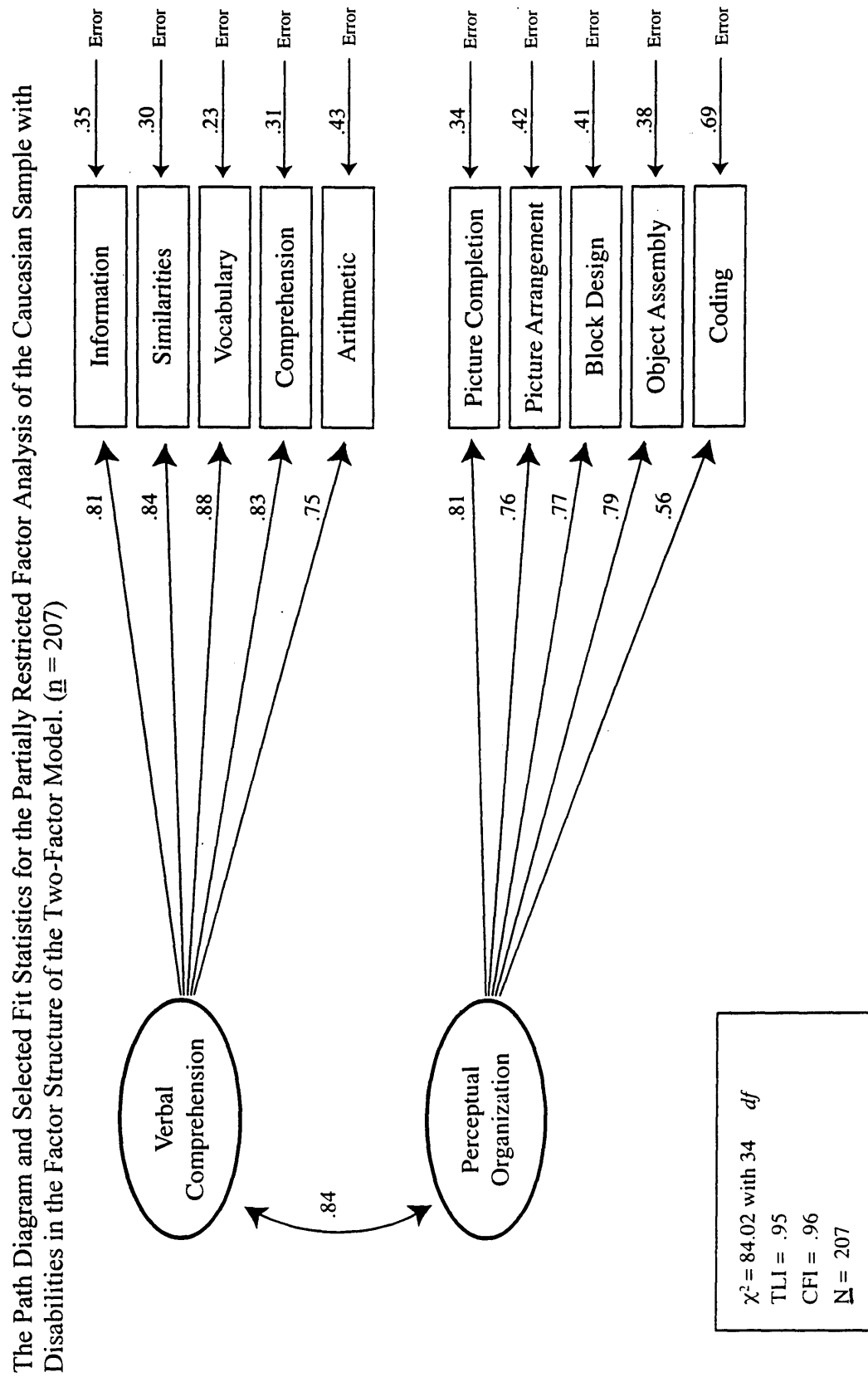
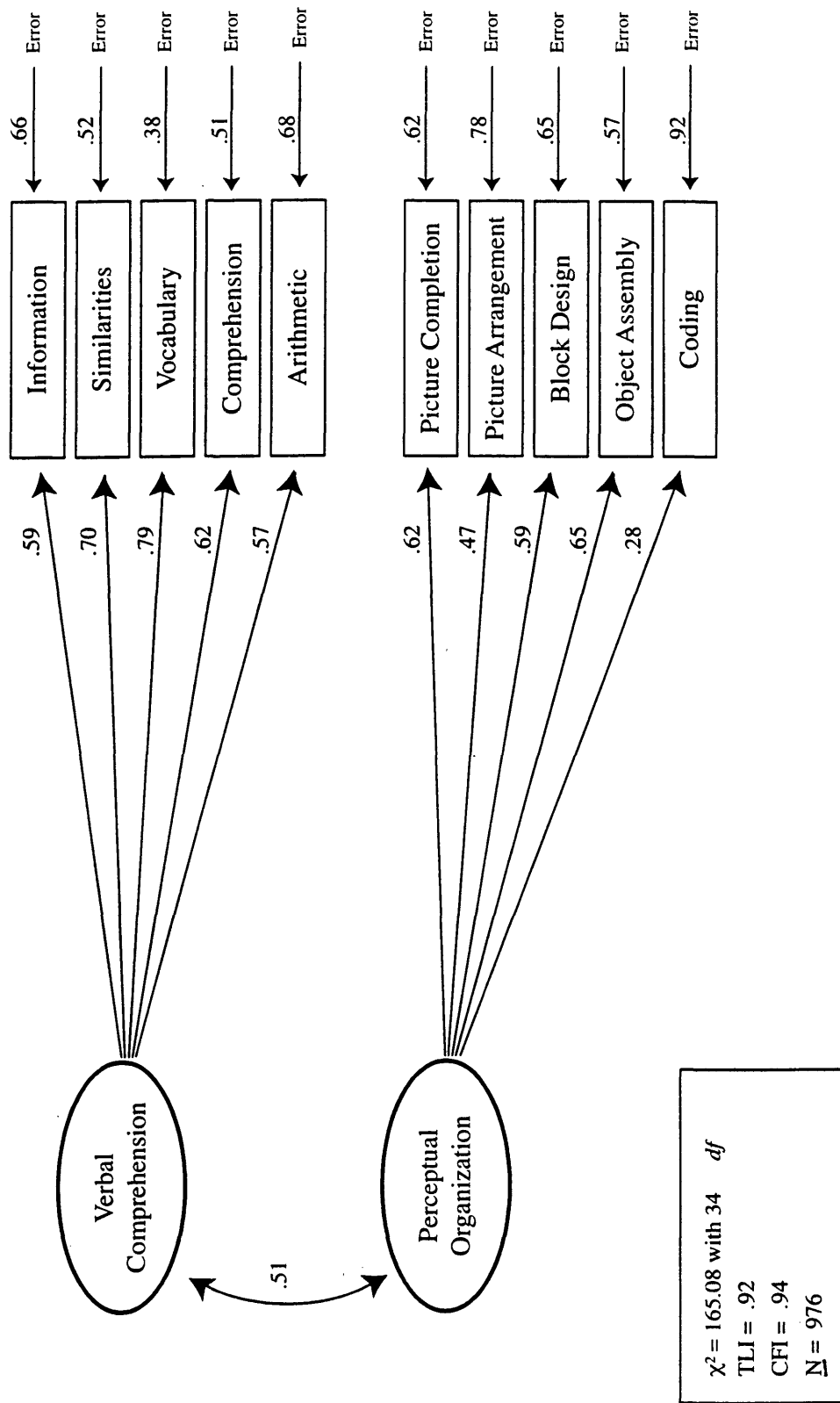


Figure 6

The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample with Learning Disabilities in the Factor Structure of the Two-Factor Model. ($n = 976$)



When the factor structure was applied back upon itself for the CN sample, the analysis, AAN sample, and the CD sample, to provide a baseline for cross-validation adequate fit was found. This is reflected in TLI and CFI values above .90 for every sample and average off diagonal standardized residual values under .08 for every sample.

Factor loadings from this analysis were examined. They were compared to those reported from an exploratory maximum-likelihood (with varimax rotation) factor analysis specifying a two-factor solution conducted from the 12 subtests for the entire standardization sample of 2,200 children in the WISC-III manual (Wechsler, 1991). Although the methodologies of these two factor analyses differ significantly, the factor loadings for the 10 core subtests used in the partially restricted factor analysis of the CN sample do not differ appreciably from those obtained for the full standardization sample as reported in Table 13. The congruence of the factor loadings, overall, across methodologies, provides evidence of factor structure reliability. The factor loadings for the 10 core subtests used in the partially restricted factor analysis of the AAN sample do not appear to differ appreciably from those obtained for the CN sample as reported in Table 11. The overall congruence of these factor loadings for the AAN sample when compared to the CN sample provides evidence of factor structure reliability across ethnicity.

Factor loadings from the partially restricted factor analysis of the CD sample were next compared to those obtained from the CN sample of the WISC-III standardization sample as reported in Table 13. Loadings appear to be congruent for six of the 10 loadings. The factor covariance for the CD sample was higher than that of the other three samples investigated in this study. These findings suggest

differences in the pattern of interrelationships among the subtests for the CD sample when it is compared to the Caucasian sample.

Table 13

Factor Loadings for Two-Factor Solutions for All WISC-III Samples and the African American Sample with Learning Disabilities

Subtest	Factor Loadings by Samples				
	SS	CN	AAN	CD	AALD
Information	.76	.79	.75	.81	.59
Similarities	.75	.79	.77	.84	.70
Arithmetic	.56	.65	.64	.75	.57
Vocabulary	.81	.81	.83	.88	.79
Comprehension	.68	.69	.74	.83	.62
Picture Completion	.50	.60	.69	.81	.62
Coding	.39	.36	.40	.56	.28
Picture Arrangement	.43	.50	.57	.76	.47
Block Design	.72	.77	.79	.77	.59
Object Assembly	.66	.70	.67	.79	.65

Note. The factor loadings for the two-factor solutions include one sample (the total WISC-III standardization sample) generated from an exploratory factor analysis and four samples generated from partially restricted factor analysis. SS = total WISC-III standardization sample; AAN = WISC-III African American Sample Without Disabilities; CD = WISC- III Caucasian Validity With Disabilities; AALD = African American Sample With Learning Disabilities.

Factor loadings from the partially restricted factor analysis of the AALD sample were next compared to the CN sample of the WISC-III standardization sample as reported in Table 11. Loadings appear to be congruent for eight of the 10 subtests.

The factor covariance was much lower between the Verbal Comprehension and Perceptual Organization factors for the AALD sample than for the other three samples. In summary, both samples without disabilities (CN and AAN) demonstrates congruence in factor loadings and factor covariances in the two-factor model. Both samples with disabilities depart from the samples without disabilities in factor loadings and factor covariances in the two-factor model but in opposite directions. The factor loadings and factor covariance for the CD sample are consistently higher than those of the sample without disabilities. The factor loadings and factor covariance for the AALD sample, by contrast, are consistently lower than those of the samples without disabilities.

Cross Validation of the Factor Structure of the Caucasian Sample (CN) of the WISC-III Standardization Sample to the African American Sample (AAN), the Clinical Validity Sample (CD) of the WISC-III Standardization Sample, and the Sample of African American Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Two-Factor Model

The structure coefficients and the factor covariance obtained from the partially restricted factor analysis of the CN sample of 1,543 children were imposed on the covariance matrices of the AAN sample of 338 children, the CD sample of 207 children, and the AALD sample of 976 children, in a fully restricted factor analysis. Fit statistics were obtained and the Lagrange Multiplier Test (LM) was run. Targeted

fit indices are reported in Table 14. A comprehensive report of all fit indices appears in Table 1 in Appendix F.

Goodness-of-Fit Statistics

The model evaluation statistics in Table 14 reveal an acceptable fit when the factor structure of the CN sample was imposed upon the AAN sample as reflected in the chi square (88.48; with $df = 43$), the TLI (.96), the CFI (.97), and the SRMSR (.06). The lack of degradation in fit suggests the construct validity of the WISC-III for the two-factor model is group-invariant across ethnicity for these two samples of children without disabilities.

Greater stress is apparent in the imposition of the factor structure of the CN sample on both clinical samples, especially as reflected in the chi-square values (CD = 166.87 and AALD = 263.86), the SRMSR's (CD = .27 and AALD = .16), and the Average R's (CD = .27 and AALD = .14). The extent to which these may be attributed to random sampling variations cannot be determined, in part because neither clinical sample was stratified. However, another important statistic, variance, which is known for both samples, can be identified as contributing to these differences. The variance for the CD group was much larger than for the CN group, whereas the variance for the AALD group was slightly smaller than that of the CN sample. As a consequence, both the SRMSR and Average R have much higher values for the mixed CD validity sample than the AALD sample. The values are not as good as the generally acceptable .08 level for the AALD sample but more discrepant in the CD sample. The CN sample consistently underestimated the values

for the CD sample but by contrast generally overestimated the values for the AALD sample as would be expected from an examination of differences in variance.

Table 14

Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample Without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Two-Factor Model

Goodness of Fit Statistics	AAN (<u>n</u> =338)	CD (<u>n</u> =207)	AALD (<u>n</u> =976)
χ^2	88.48	166.87	263.86
<u>df</u>	43	43	43
SRMSR	.06	.27	.16
Average Residual	.05	.27	.14
TLI	.96	.90	.90
CFI	.97	.91	.90

Note. AAN = WISC-III African American Sample Without Disabilities;

CD = WISC- III Caucasian Validity With Disabilities; AALD = African American

Sample With Learning Disabilities; SRMSR = standardized root mean square

residual; Average R = average off-diagonal absolute standardized residual;

TLI = Tucker-Lewis index; CFI = comparative fit index.

The TLI and CFI are marginally acceptable for both clinical samples, hovering about the .90 value. Overall, they suggest the two-factor model generalizes reasonably well from the CN sample without disabilities to both samples with disabilities, providing evidence of the invariance of the factor structure of the WISC-III for the two-factor model.

LM Test

The results of the cumulative multivariate statistics LM test for both clinical samples were reviewed. Because the results of the cumulative multivariate LM test statistics take into account restrictions in the model and correlations, it is more judicious to base respecifications on them (Byrne, 1994b). The inspection of the cumulative multivariate LM test statistics revealed the same major malfitting parameter in both clinical samples. The LM test statistic specifies freeing the structural path involving the Verbal factor and the Performance factor covariance. Freeing the structural path for factor covariance does not violate the basic assumptions of the two-factor model in terms of specifying which subtests load on which factor. Rather, it provides data that addresses a different question of theoretical importance about the possible relationship of the differentiation of human abilities to the cognitive maturation process (Garrett, 1946; Quereshi, 1967) in these two samples with disabilities.

When freeing the structural path for the factor covariance was allowed, the improvement in the chi-square statistic was 56 points with regard to the CD sample and 49 points with regard to the AALD sample. Fit statistics and residuals for the respecified clinical samples are reported in Table 15.

When the respecification was accomplished, the improvement in all model evaluation statistics for both samples with disabilities was noteworthy. The TLI and CFI goodness-of-fit values were more robust, reflecting an acceptable fit. The reproduced correlations for the respecified two-factor model resulted in far less misspecification. The SRMSR for the AALD sample was reduced to .07 from an initial value of .16. This new value was well within acceptable limits and was nearly

identical to the SRMSR for the AAN sample in the initial analysis. The improvement in the SRMSR for the CD sample from an initial value of .27 to a new value of .11 was even greater, reflecting a much better fit, albeit one still reflecting greater misspecification. The interfactor correlations in the respecified factor analyses are as follows: .83 for the CD sample; and .48 for the AALD sample. Evidence for the generalizability of the factor structure from the WISC-III Caucasian sample without disabilities to both samples differing in ethnicity and disability status was stronger in the respecified two-factor model.

Table 15

Model Evaluation Statistics for the Respecified Factor Structure and Loadings of the Caucasian Sample without Disabilities Imposed on the Caucasian Sample with Disabilities and the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Two-Factor Model

Goodness of Fit Statistics	CD ($n = 207$)	AALD ($n = 976$)
χ^2	124.35	233.63
<u>df</u>	42	42
SRMSR	.11	.07
Average R	.09	.06
TLI	.93	.92
CFI	.94	.93

Note. CD = WISC- III Caucasian Validity With Disabilities; AALD = African American Sample With Learning Disabilities; SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

Cross-Validation of the Factor Structure of the African American Sample (AAN) of
the WISC-III Standardization Sample to the African American Sample of Children
with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the
Two-Factor Model

The factor loadings and factor covariance obtained from the partially restricted factor analysis of the AAN sample of 338 children were imposed on the covariance matrix of the AALD validity sample of 976 children in a fully restricted factor analysis. Fit statistics were obtained and results related to fit statistics were presented in Table 16. The Lagrange Multiplier Test (LM) was also conducted. A comprehensive report of all fit indices appears in Table 2 in Appendix F.

Goodness-of-Fit Statistics

As reported in Table 16, the TLI (.89) and CFI (.90) reflected a marginally acceptable fit to the data in this cross-validation application. They did not differ appreciably from the values obtained in the cross-validation with the CN sample. The values of the SRMSR and Average R reflected some degree of misfit in the model of the AAN loadings to the AALD sample. The extent to which this is related to random sampling fluctuations cannot be determined. The relationship of the degree of misspecification to differences in variance can be explored, however. The variance for the AALD sample was smaller than the variance for the AAN sample. This difference could affect the degree of error in specification, causing the loadings from the AAN sample to overestimate the observed covariances. The values of the SRMSR and the Average R did not differ appreciably from the values obtained in the cross-validation with the CN sample, revealing about the same degree of specification error

in the two-factor model across both ethnicity samples without disabilities to the African American sample with learning disabilities.

Table 16

Model Evaluation Statistics for the Factor Structure and Loadings of the African American Sample without Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Two-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AALD</u>
χ^2	269.28
<u>df</u>	43
SRMSR	.16
Average R	.14
TLI	.89
CFI	.90

Note. AALD = African American Sample with Learning Disabilities;

SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

LM Test

Inspection of the cumulative multivariate LM test statistics along with their accompanying univariate increments for the fully restricted factor analysis of the AALD sample followed next. It revealed the same major malfitting parameter specified when cross-validation was accomplished involving imposition of the factor structure of the CN sample on the CD and AALD clinical samples. The LM test statistic specified freeing up the structural path involving the Verbal factor and

Performance factor covariance. When this was allowed, the improvement in the chi-square was 53 points. Fit statistics and residuals for the respecified model are reported in Table 17.

Table 17

Model Evaluation Statistics for the Respecified Factor Structure and Loadings of the African American Sample without Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Two-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AALD (n = 976)</u>
χ^2	236.26
<u>df</u>	42
SRMSR	.07
Average R	.06
TLI	.92
CFI	.93

Note. AALD = African American Sample With Learning Disabilities; SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

When the respecification was reviewed it was apparent there was improvement in all model evaluation statistics. The improvement in the TLI and CFI were such that an acceptable fit was achieved (as reported in Figure 8). More importantly, however, the reproduced correlations for the respecified two-factor model moved well within acceptable limits, representing a considerable reduction in misspecification (i.e., from .16 for the SRMSR to .07 and from .14 for the Average R

to .06). The interfactor correlation for the AALD sample in respecification was .50. The two-factor model in respecification provided stronger evidence of the generalizability of the factor structure within ethnicity from the African American standardization sample without disabilities to the African American sample of children with learning disabilities.

Cross-Validation of the Factor Structure of the Caucasian Validity Sample (CD) of the WISC-III Standardization Sample to the African American Sample of Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Two-Factor Model

The factor loadings and factor covariance obtained from the partially restricted factor analysis of the CD sample of 207 children were imposed on the covariance matrix of the AALD validity sample of 976 children in a fully restricted factor analysis. Fit statistics were obtained and results related to fit statistics are presented in Table 18. A comprehensive report of all fit indices for the two-factor model appears in Table 3 in Appendix F.

Goodness-of-Fit Statistics

As reported in Table 18, the TLI (.75) and CFI (.77) reflected a poor fit to the data. The consistently poor statistics obtained with respect to all goodness of fit indices suggests considerable misspecification in the model of the CD loadings to the AALD sample, reflected in the Average R value of .57 and the SRMR value of .59. All residuals have negative values, suggesting the CD loadings consistently overestimate the values for the AALD sample. A review of the standardized residuals revealed an extreme shift of the distribution in the negative direction.

Table 18

Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample with Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Two-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AALD</u>
χ^2	566.81
<u>df</u>	43
SRMSR	.59
Average R	.57
TLI	.75
CFI	.77

Note. AALD = African American Sample with Learning Disabilities;

SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

LM Test

The Lagrange Multiplier Test (LM) was also conducted. Inspection of the cumulative multivariate LM test statistics along with their accompanying univariate increments revealed the same major misfitting parameter reported in the samples with disabilities previously. The major parameter involved freeing the structural path involving the Verbal factor and Performance factor covariance. When this was allowed the improvement in the chi-square was 199 points. Fit statistics and residuals for the respecified AALD sample are reported in Table 19.

Table 19

Model Evaluation Statistics for the Respecified Factor Structure and Loadings of the Caucasian Sample with Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Two-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AALD</u>
χ^2	238.28
<u>df</u>	42
SRMSR	.06
Average R	.05
TLI	.92
CFI	.93

Note. AALD = African American Sample with Learning Disabilities;

SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

The examination of the respecified values (as reported in Figure 8), including the TLI (.92), CFI (.93), and RSMR (.06) was conducted. It revealed the greatest improvement in model evaluation statistics across all of the analyses when the factor covariance was freed up and allowed to be estimated in the imposition of the factor structure of the CD sample on the AALD sample. The interfactor correlation in the respecified analysis for the AALD sample was .52. The two-factor model fit was very good overall, except with respect to factor covariances in these two samples with disabilities. This suggests that the degree to which ability factors covaried in this sample of Caucasian school-age children with mixed disabilities was very different

from the sample of African American children with learning disabilities. Both samples with disabilities differed in the amount of covariance for which ability factors accounted when compared to the two samples without disabilities. Disability status, rather than ethnicity, emerged as the variable that was associated with greater model misspecification.

Overview of Results for the Two-Factor Model

Assessment of Overall Model Fit

Figure 7, entitled “Overview Model Fit of Two-Factor Model Across Samples,” illustrates the overall model fit of the two-factor model across the four samples prior to respecification. The critical nature of the imposition of the factor loadings from the Caucasian samples initially upon all other samples is emphasized by its position at the apex of the figure and its application upon the three samples represented below it. A unidirectional arrow indicates the imposition of the factor loadings from one sample upon another in a fully restricted factor analysis in every analysis. The tail of the arrow emanates from the sample where factor loadings are being imposed upon another sample. The head of the arrow rests against the rectangle representing the sample upon which the factor loadings have been imposed. The TLI and CFI values presented in proximity to the arrow heads represent the fit statistics in the fully restricted factor analysis for the two-factor model. The TLI and CFI scores that are enclosed in boxes reflect values for the partially restricted factor analyses performed previously upon each sample. These analyses had generated the factor loadings imposed in the fully restricted factor analyses illustrated in this figure.

The TLI and CFI for samples with disabilities are reported in the respecified two-factor overview in Figure 8. Step-wise, initially moving from the apex of the diagram representing the Caucasian sample without disabilities, arrows descend to the Caucasian sample with disabilities and the African American sample with learning disabilities accompanied by TLI and CFI values in respecification. At the second level, on the left side of the figure, the arrow from the African American sample without disabilities descends to the African American sample with learning disabilities accompanied by TLI and CFI values in respecification. Lastly, the arrow from the Caucasian sample with disabilities descends to the African American sample with learning disabilities accompanied by TLI and CFI values in respecification. All of these values reflected adequate fit in respecification.

Figure 7

Overall Model fit of Two-Factor Model Across Samples

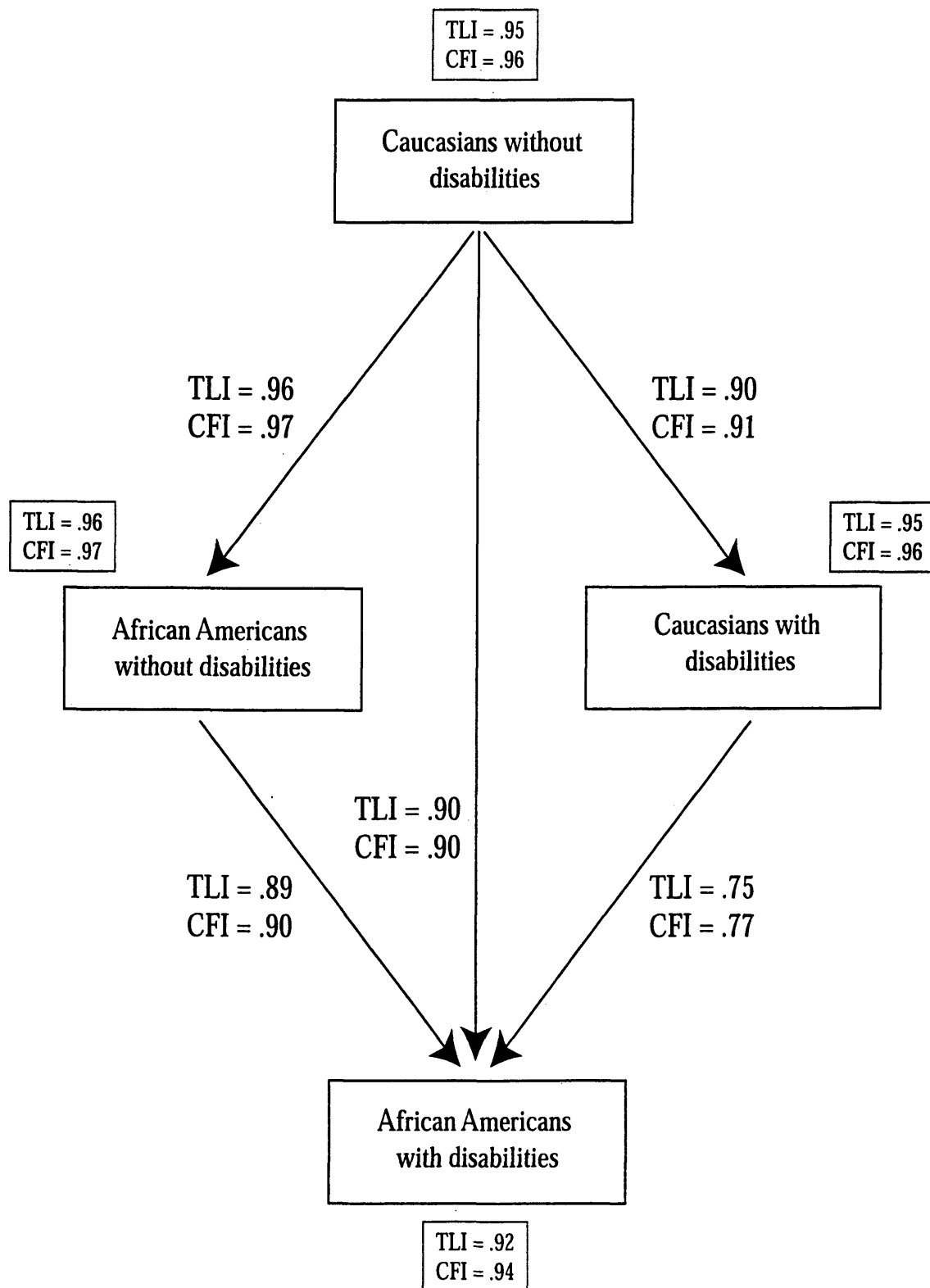
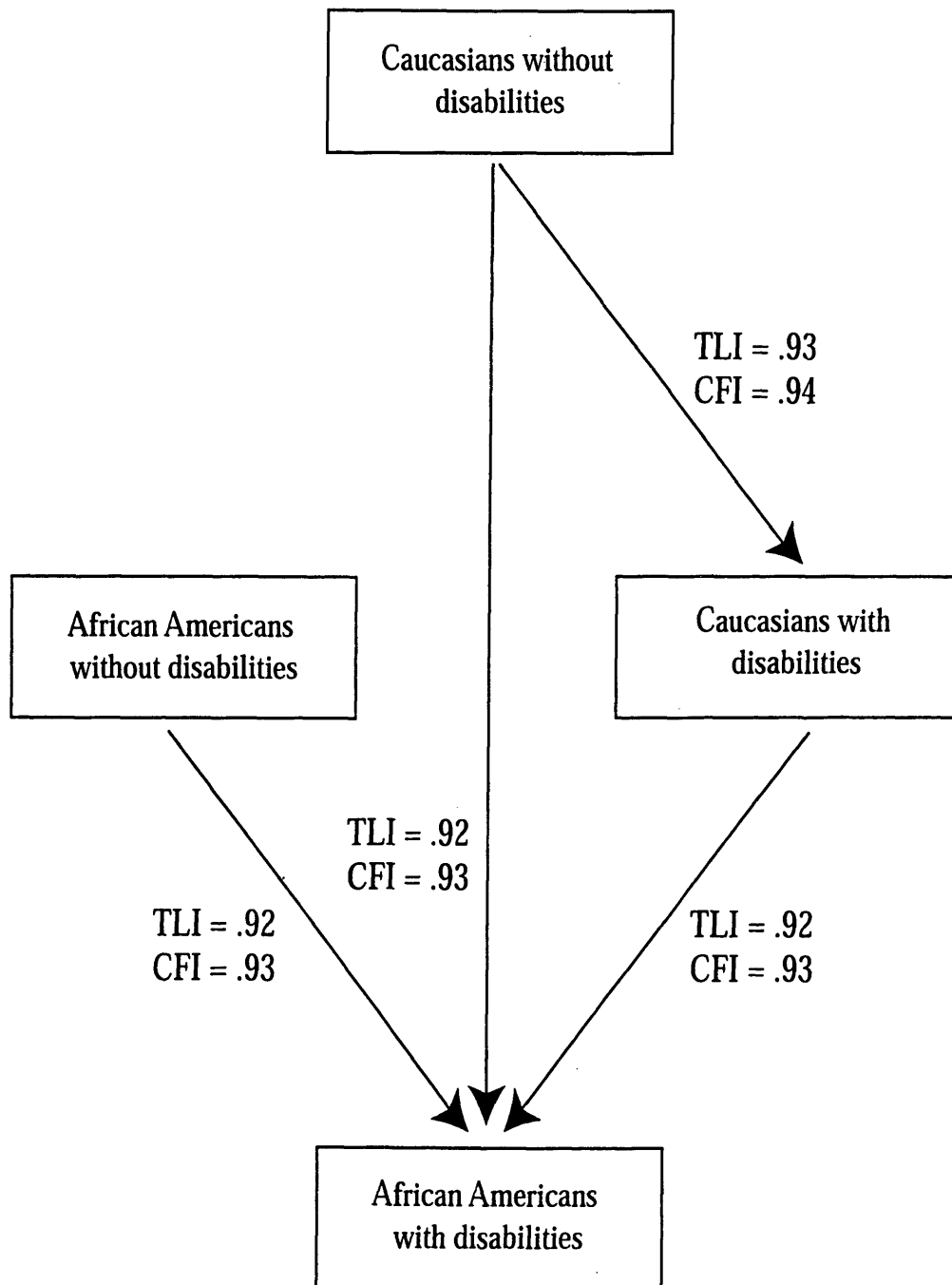


Figure 8

Respecified Model Fit of the Two-Factor Model Across Samples



Investigation of the Three-Factor Model

Overview

The ten core subtests and one supplementary subtest of the WISC-III corresponding to the theoretical model of the three-factor model are: Information, Similarities, Vocabulary, Comprehension, Arithmetic, Digit Span, Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding. The path diagram representing the theoretical structure of the three-factor model of the WISC-III is reported in Figure 9.

Partially Restricted Factor Analyses of the Caucasian Sample (CN), the African American Sample (AAN) of the WISC-III Standardization Sample, the Caucasian Validity Sample (CD) and the African American Sample of Children with Learning Disabilities (AALD) for the Three-Factor Model

The structure coefficients for the subtests and the factor covariances for the four samples are reported in Table 20. Model evaluation statistics for the four samples are reported in Table 21 for the three-factor model. The path diagrams for the three-factor model for each sample including factor loadings, factor covariances, error variances of the subtests, and selected fit statistics are contained in Figures 10-13. When the factor structure was applied back upon itself for the CN sample, the AAN sample, the CD sample, and the AALD sample, to provide a baseline for cross-validation adequate fit was found as reported in Table 21. This is reflected in TLI and CFI values above .90 for every sample and average off diagonal standardized residual values under .08 for every sample.

Figure 9

Hypothesized First-Order CFA Model for the Three-Factor Model

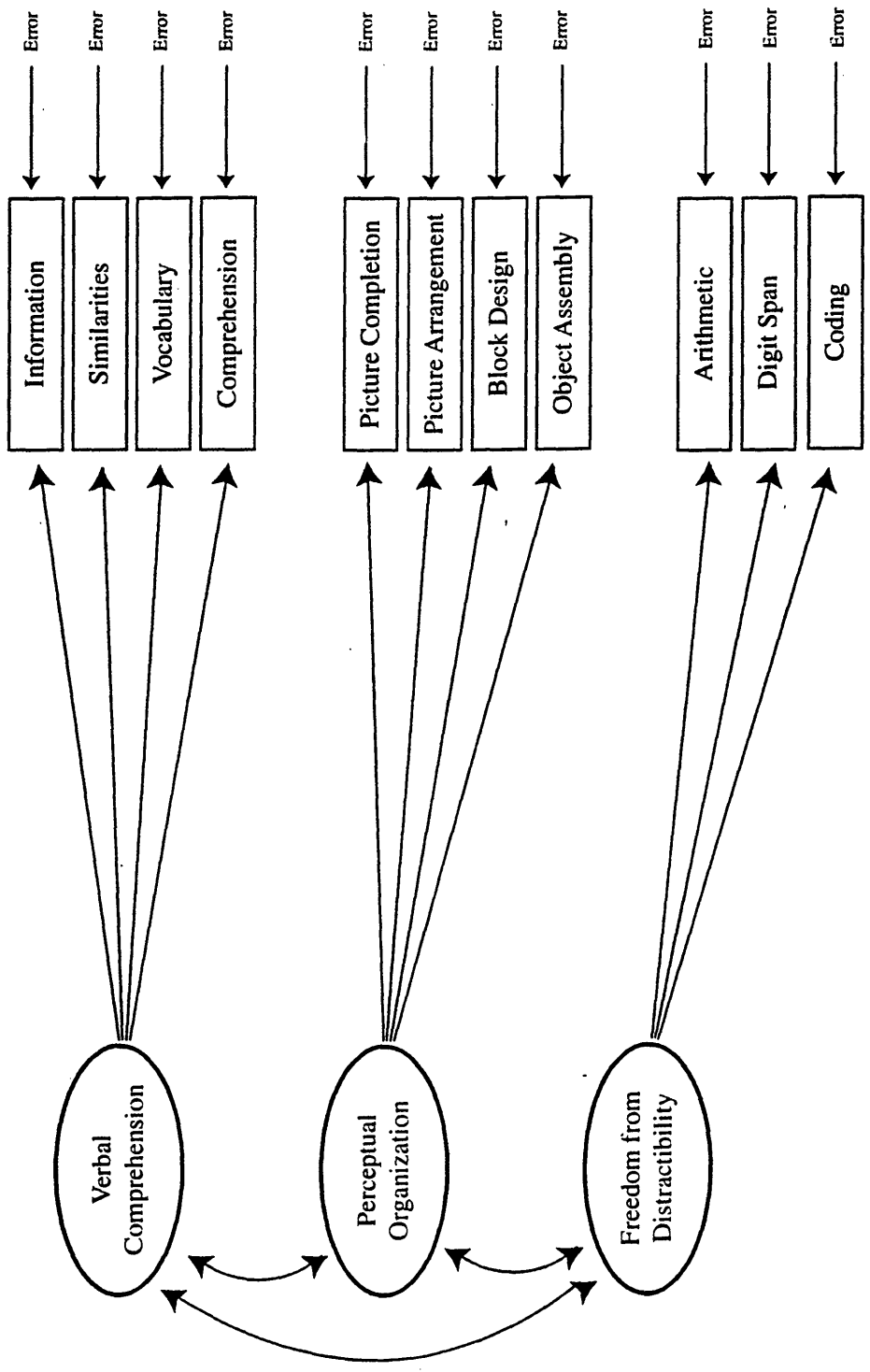


Table 20

Partially Restricted Factor Analysis for the Three-Factor Model

	<u>CN</u>	<u>AAN</u>	<u>CD</u>	<u>AALD</u>
Subtests	(<u>n</u> = 1,543)	(<u>n</u> = 338)	(<u>n</u> = 207)	(<u>n</u> = 646)
<u>Factor 1 Verbal Comprehension Path Coefficient</u>				
Information	.79	.75	.81	.54
Similarities	.79	.77	.85	.66
Vocabulary	.82	.83	.88	.77
Comprehension	.69	.79	.84	.69
<u>Factor 2 Perceptual Organization Path Coefficient</u>				
Picture Completion	.60	.70	.81	.64
Picture Arrangement	.49	.56	.76	.42
Block Design	.79	.79	.78	.55
Object Assembly	.70	.67	.78	.65
<u>Factor 3 Freedom From Distractibility Path Coefficient</u>				
Arithmetic	.80	.73	.82	.85
Digit Span	.52	.59	.72	.36
Coding	.36	.43	.57	.21
<u>Factor Covariances</u>				
F ₁ , F ₂	.69	.75	.82	.50
F ₂ , F ₃	.72	.76	.92	.47
F ₁ , F ₃	.77	.83	.86	.65

Note. Structure coefficients are reported for each sample in each factor for the three-factor model as well as factor covariances for each sample in the three-factor model.

Table 21

Model Evaluation Statistics for the Partially Restricted Factor Analysis for the Three-Factor Model

<u>Subtests</u>	<u>Goodness-of-fit statistics</u>				
	χ^2	<u>df</u>	Average R	TLI	CFI
WISC-III Caucasian subset (<u>n</u> = 1,543)	227.05	41	.03	.96	.97
WISC-III African American subset (<u>n</u> = 338)	83.67	41	.03	.96	.97
WISC-III Caucasian subset with Disabilities (<u>n</u> = 207)	78.75	41	.03	.97	.97
African American Sample with Learning disabilities (<u>n</u> = 646)	125.36	41	.04	.92	.94

Note. TLI = Tucker-Lewis index; CFI = comparative fit index; Average R = Average off-diagonal absolute standardized residual.

A comparison of the TLI and CFI obtained for the CN sample, and CD sample for the three-factor model (as seen in Table 21) to those obtained for the two-factor model (as seen in Table 12) revealed trivial differences. Factor loadings obtained from these analyses are reported in Figures 10, 11, 12, and 13. They were compared to those for subtests from the partially restricted factor analyses reported for the two-factor model in Figures 3, 4, 5, and 6. Only trivial differences were noted except for Arithmetic. The loading for Arithmetic consistently was larger in the Freedom from

Figure 10

The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the Caucasian Sample without Disabilities in the Factor Structure of the Three-Factor Model. ($n = 1,543$)

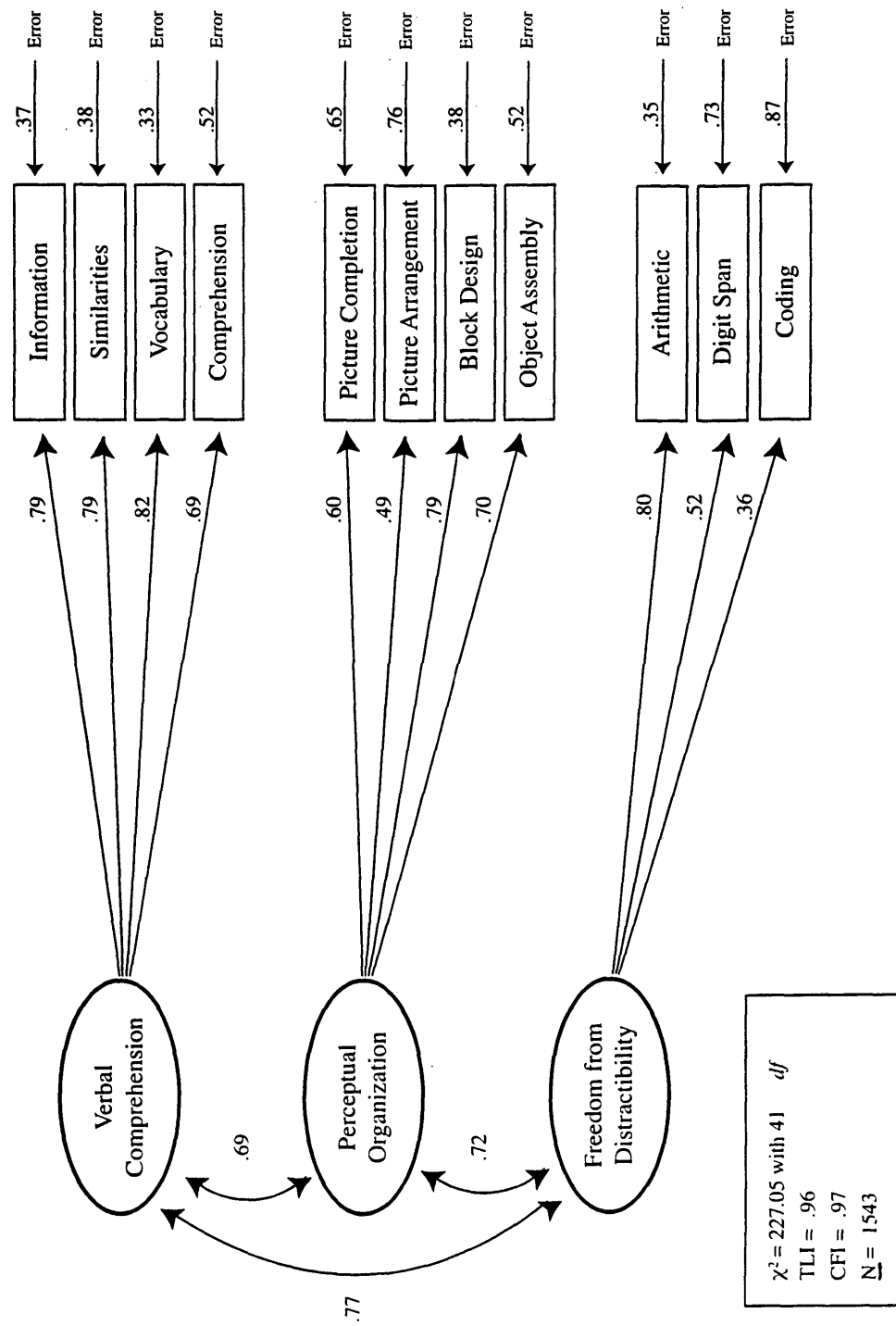


Figure 11

The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample without Disabilities in the Factor Structure of the Three-Factor Model. ($n = 338$)

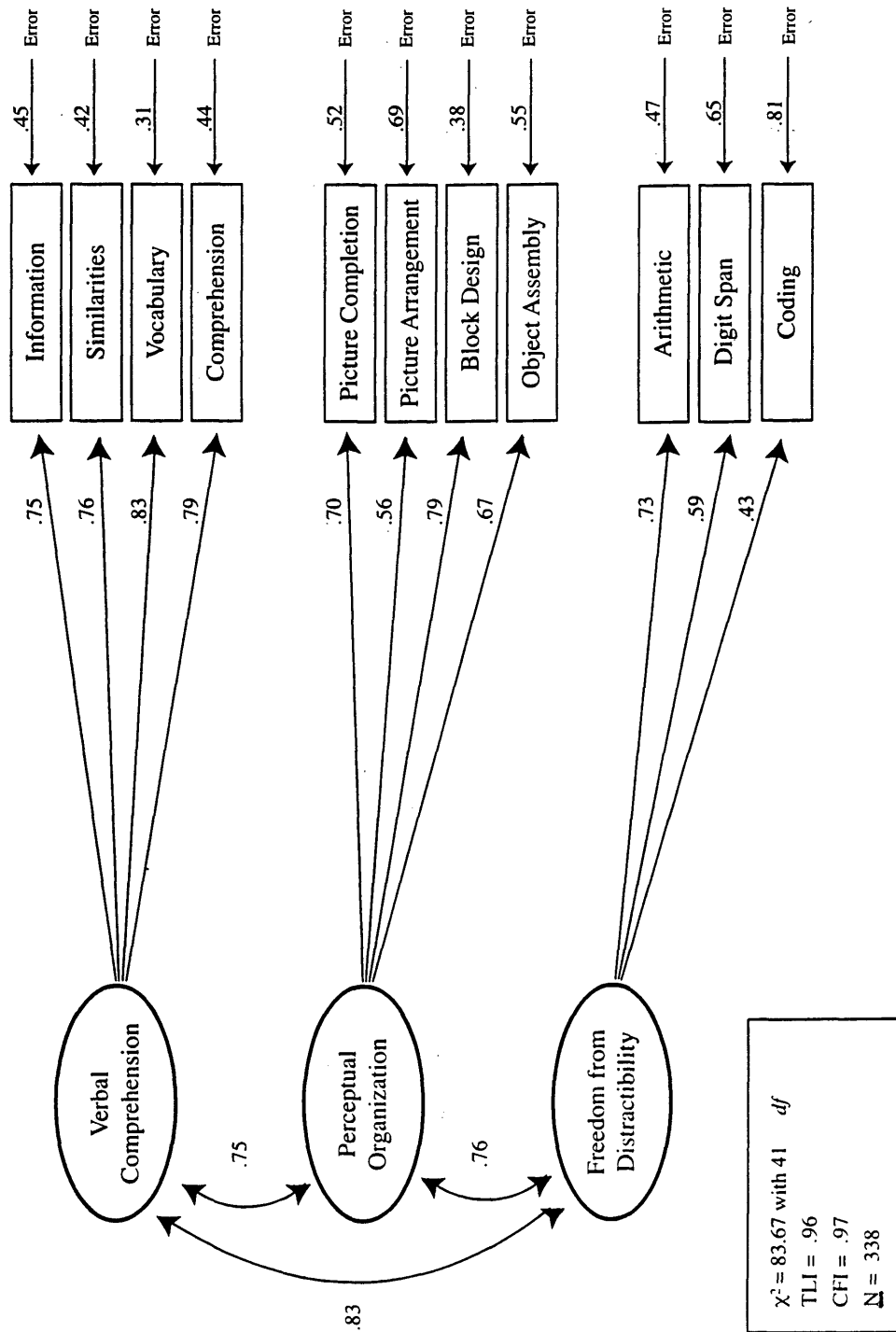


Figure 12

The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the Caucasian Sample with Disabilities in the Factor Structure of the Three-Factor Model. ($n = 207$)

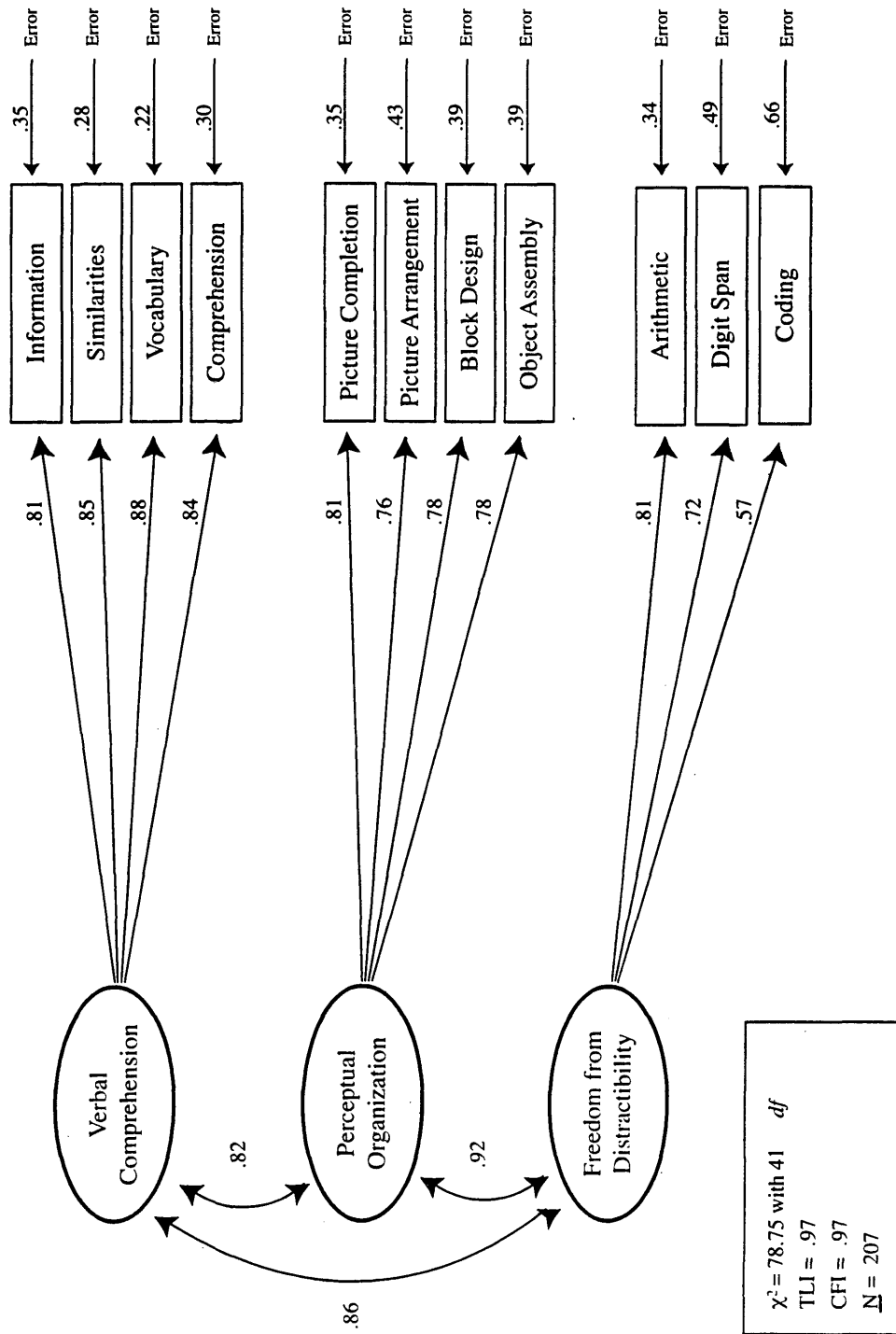
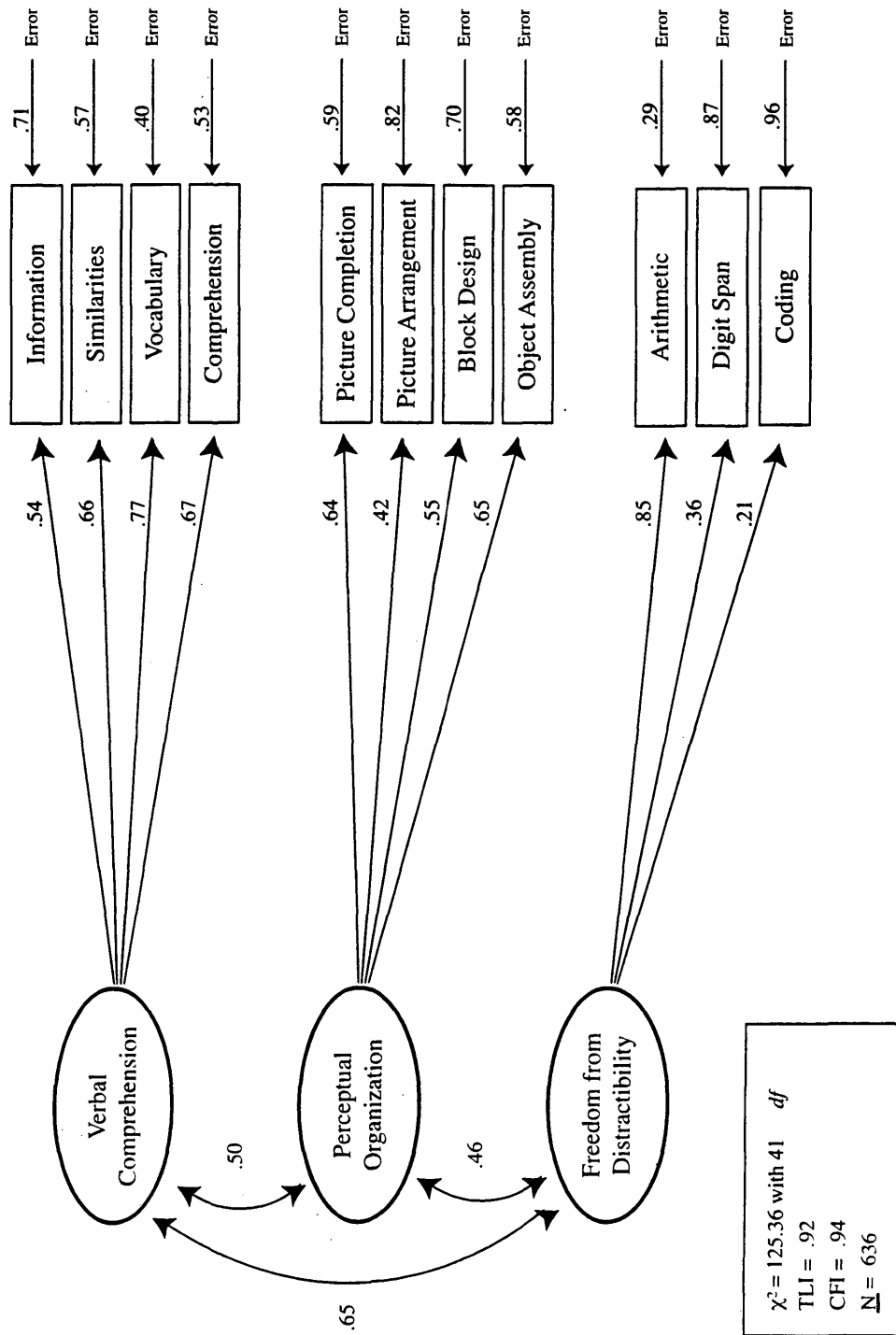


Figure 13

The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample with Learning Disabilities in the Factor Structure of the Three-Factor Model. ($n = 646$)



Distractibility factor than it had been on the Verbal Comprehension factor. This change was congruent with statistical procedures used in specifying an additional factor in the initial factor solution for the three-factor model and allowing Arithmetic to join with Digit Span and Coding to form that factor.

Cross Validation of the Factor Structure of the Caucasian Sample (CN) of the WISC-III Standardization Sample to the African American Sample (AAN), the Clinical Validity Sample (CD) of the WISC-III Standardization Sample, and the Sample of African American Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Three-Factor Model

The factor loadings and factor covariance obtained from the partially restricted factor analysis of the CN sample of 1,543 children were imposed on the covariance matrices of the AAN sample of 338 children in a fully restricted factor analysis. Additionally they were imposed on the CD sample of 207 children and the AALD sample of 646 children, in a fully restricted factor analysis. Fit statistics were obtained and the Lagrange Multiplier Test (LM) was run. Targeted fit indices are reported in Table 22. A comprehensive report of all fit indices for the three-factor model appears in Table 4 in Appendix F.

The model evaluation statistics in Table 22 suggest an excellent fit when the factor structure of the CN sample imposed on the AAN sample, as reflected in the chi square (114.84), the TLI (.95), the CFI (.96), and the SRMSR = .08. The adequacy of the fit suggests the construct of the WISC-III for the three-factor model is group invariant across ethnicity in these two samples of children without disabilities.

Table 22

Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Three-Factor Model

Goodness of Fit Statistics	AAN ($n = 338$)	CD ($n = 207$)	AALD ($n = 646$)
χ^2	114.84	178.25	200.20
<u>df</u>	52	52	52
SRMSR	.08	.29	.18
Average R	.06	.29	.15
TLI	.95	.91	.88
CFI	.96	.91	.89

Note. AAN = WISC-III African American Sample Without Disabilities;

CD = WISC- III Caucasian Validity Sample With Disabilities; AALD = African American Sample With Learning Disabilities; SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

The issue of how to compare the fit of competing models statistically emerges. Thorndike (1978) cogently elucidated a rationale for comparison of competing models.

The first and foremost criterion of adequacy for a set of factors is the accuracy with which they reproduce the correlations for which they were developed. A potentially useful way to judge whether an additional factor adds enough information beyond that already contained in a set of factors is to compare the matrix of residual correlations obtained with the extra factor to a matrix obtained without it. Because the residuals are the differences between the original and the reproduced correlations, the addition of a useful factor to the

set should result in a noticeable decrease in the size of these residuals (Thorndike, p.270).

Inspection of the residuals for the two-factor and three-factor models reveals an increase in the SRMSR from .06 for the two-factor model to .08 for the three-factor model. The Average Residual for the two-factor model is .05 and for the three-factor model .06. The increase in the size of the residuals from the two-factor model to the three-factor model may have been due to fixing the two extra factor correlations.

As was reported for the two-factor model greater stress was again apparent in the imposition of the factor structure of the CN sample on both clinical samples for the three-factor model. This was especially reflected in the chi-square values (178.25 for the CD sample and 200.20 for the AALD sample, $df = 52$), the SRMSR's, (CD sample = .29 and AALD = .18), and the Average R's (CD = .29 and AALD = .15). The extent to which these reflect random sampling variation cannot be determined for reasons stated previously. Only a portion of the AALD sample ($n = 676$) were administered the eleventh subtest, Digit Span.

Once again, it was noted the SRMSR and Average R had even higher values for the CD sample than the AALD sample. The misspecification for both samples was slightly larger than it was for the two-factor model. The largest standardized residuals for the AALD sample in the CN factor loadings are reported in Table 2 in Appendix G. About 68% of the distribution of residuals for the AALD sample falls between -.3 and -.1, with an additional 3% falling between -.3 and -.4. Only 29% fall in the higher ranges from .0 to .1. This suggests restriction of range was responsible for some misspecification in the AALD sample. The slight increase in the size of the

residuals for the CD sample congruent with increases reported for the AALD sample, suggested the three factor solution did not reproduce the correlations more accurately than the two-factor model.

The values of the TLI (.91) and CFI (.91) for the CD sample suggested the three-factor solution provided a marginally acceptable fit. Values of these fit indices were trivially higher than they were for the two-factor model. The TLI (.88) and CFI (.89) were trivially lower for the AALD sample for the three-factor model than the two-factor model. These variations probably reflected sampling variations. Results of the cumulative multivariate LM statistics revealed no changes that could be made without seriously violating the assumptions of the three-factor model. In contrast to the two-factor model, there was no indication that freeing the factor covariances would result in significant improvement in overall model fit.

Cross-Validation of the Factor Structure of the African American Sample (AAN) of
the WISC-III Standardization Sample to the African American Sample of Children
with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the
Three-Factor Model

The factor loadings and factor covariances obtained from the partially restricted factor analysis of the AAN sample of 338 children were imposed on the covariance matrix of the AALD validity sample of 636 children in a fully restricted factor analysis. Fit statistics were obtained and results related to the fit statistics are presented in Table 23.

Table 23

Model Evaluation Statistics for the Factor Structure and Loadings of the African American Sample without Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Three-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AALD</u>
χ^2	214.20
<u>df</u>	52
SRMSR	.17
Average R	.16
TLI	.87
CFI	.88

Note. AALD = African American Sample with Learning Disabilities;

SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

The rather large chi square value (214.20 with df = 52), the marginally unacceptable values of the TLI (.87), and the CFI (.88), and the elevated values of the SRMSR (.17) and Average R (.15) suggest greater misspecification in the three-factor model than the two-factor model. When the distribution of the standardized residuals was examined, as reported in Table 2 in Appendix F, it was apparent that the factor loadings from the reproduced correlations overestimated the value of the intercorrelations to a greater extent in the three-factor model than the two-factor

model. Differences in variances of some of the subtests may account for some of the discrepancy.

The inspection of the cumulative multivariate LM test statistics along with their accompanying univariate increments revealed only modifications that would violate the assumptions of the three-factor model seriously. Individually, none of the modifications improved the chi-square substantially. This suggested the model fit in general was not good.

Cross-Validation of the Factor Structure of the Caucasian Validity Sample (CD) of the WISC-III Standardization Sample with Disabilities, to the African American Sample of Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Three-Factor Model

The factor loadings and factor covariance obtained from the partially restricted factor analysis of the CD sample of 207 children were imposed on the covariance matrix of the AALD validity sample of 636 children in a fully restricted factor analysis. Fit statistics were obtained and results related to fit statistics are presented in Table 24. A comprehensive report of all fit indices appears in Table 2 in Appendix F.

The fit statistics for the three-factor model, including the rather large chi-square (414.29 with $df = 52$), the low values of the TLI (.72) and the CFI (.73), and the large values of the SRMSR (.59) and Average R (.56) all suggested there was considerable misspecification in the model of the CD loadings to the AALD sample. The inspection of the residuals (reported in Table 3 in Appendix G) consistent with statistics for the two-factor model, revealed that all residuals had negative values,

suggesting the CD loadings consistently overestimated the values for the AALD sample. Wide differences in variance for the subtest scores between the two samples may account for much of the error. That the model fit was generally bad is reflected in the results of the LM test that produces only suggestions for modifications in parameters that seriously violated assumptions about the model and did not substantially improve the chi-square statistic.

Table 24

Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample with Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Three-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AALD</u>
χ^2	414.29
<u>df</u>	52
SRMSR	.59
Average R	.56
TLI	.72
CFI	.73

Note. AALD = African American Sample with Learning Disabilities;

SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

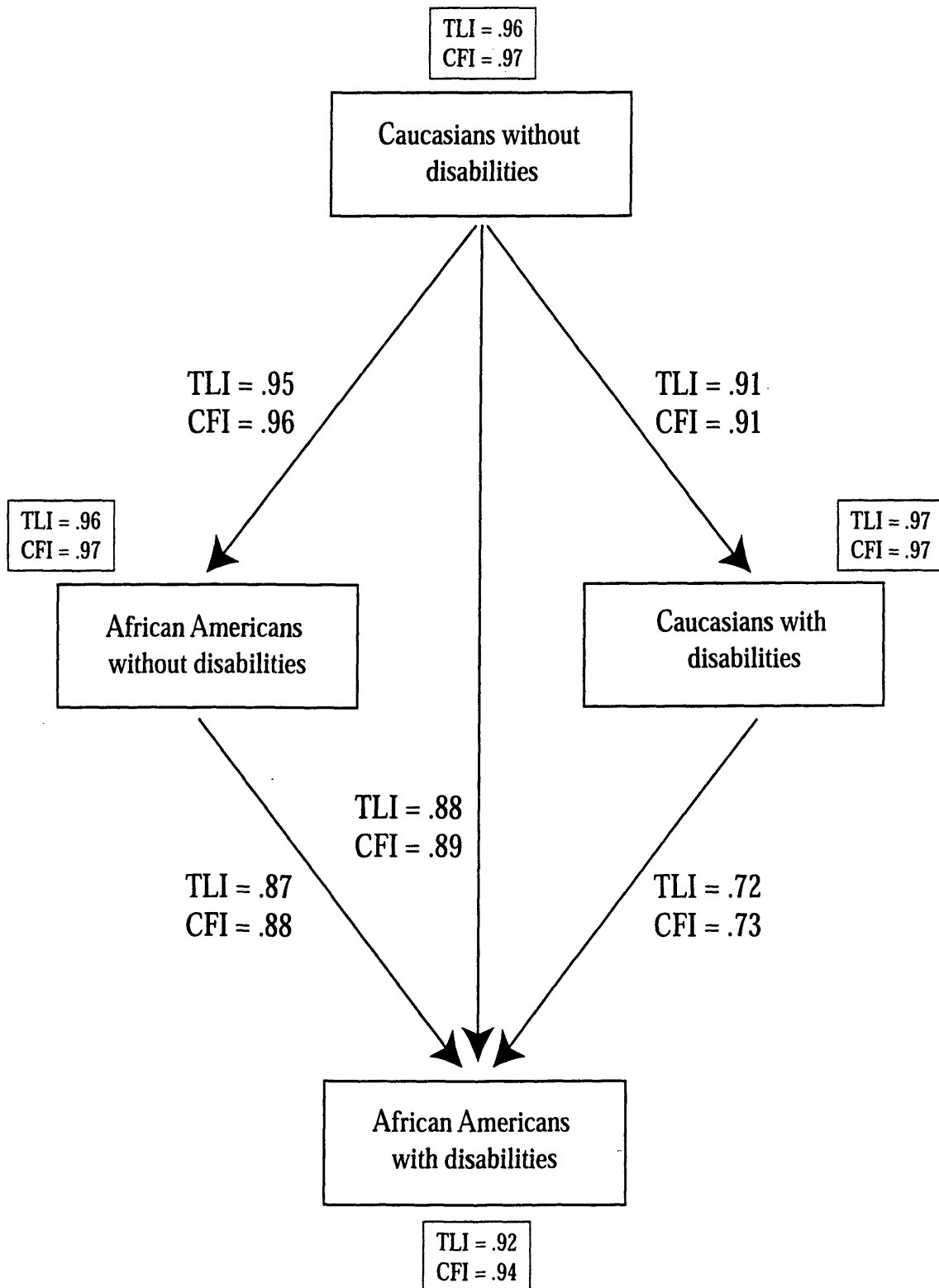
Overview of Results for the Three-Factor Model

Assessment of Overall Model Fit

Figure 14 entitled “Overview of Model Fit of Three-Factor Model Across Samples” illustrates the overall model fit of the three-factor model across the four samples. The critical nature of the imposition of the factor loadings from the Caucasian samples initially upon all other samples is emphasized by its position at the apex of the figure as its application upon the three samples represented below it. The imposition of the factor loadings from one sample upon another in a fully restricted factor analysis in every analysis is indicated by a unidirectional arrow. The tail of the arrow emanates from the sample where factor loadings are being imposed upon another sample. The head of the arrow rests against the rectangle representing the sample upon which the factor loadings have been imposed. The TLI and CFI values presented in proximity to the arrow heads represent the fit statistics in the fully restricted factor analysis for the four-factor model. The TLI and CFI scores that are enclosed in boxes reflect values for the partially restricted factor analyses performed previously upon each sample. These analyses had generated the factor loadings imposed in the fully restricted factor analyses illustrated in this figure.

Figure 14

Overall Model Fit of Three-Factor Model across Samples.



Investigation of the Four-Factor Model

Overview

The ten core subtests and two supplementary subtests of the WISC-III corresponding to the theoretical model of the four-factor model are: Information, Similarities, Vocabulary, Comprehension, Arithmetic, Digit Span, Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, and Symbol Search. The path diagram representing the theoretical structure of the four-factor model of the WISC-III is reported in Figure 15.

Partially Restricted Factor Analysis of the Caucasian Sample (CN), the African American Sample (AAN) of the WISC-III Standardization Sample, the Caucasian Validity Sample (CD) and the African American Sample of Children with Learning Disabilities (AALD) for the Four-Factor Model

Initially, a partially restricted maximum-likelihood factor analysis was conducted of the WISC-III CN sample of 1,543 children, the African American sample (AAN) of 338 children, the Caucasian validity sample (CD) of 207 children with disabilities, and the African American sample of 172 children with learning disabilities (AALD) for whom all 12 subtests were available. The covariance matrices generated from the 12 core subtests specified in the preceding section were used in the partially restricted factor analysis. The structure coefficients for the subtests and the factor covariance for the four samples are reported in Table 25. Model evaluation statistics for the four samples are reported in Table 26 for the four-factor model. A comprehensive report of all fit indices appears in Table 5 in

Figure 15

Hypothesized First-Order CFA Model for the Four-Factor Model.

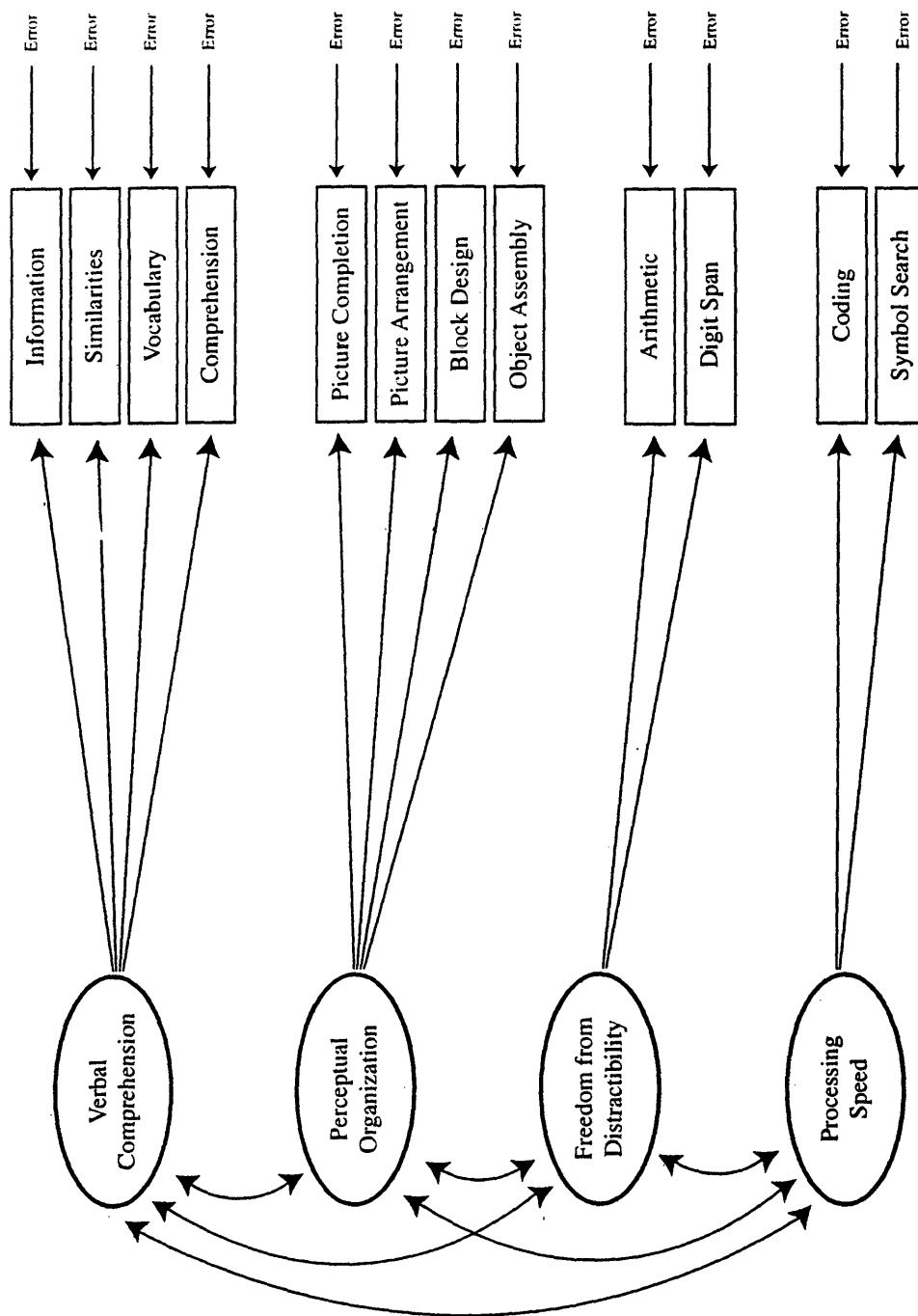


Table 25

Partially Restricted Factor Analysis for the Four-Factor Model

	<u>CN</u>	<u>AAN</u>	<u>CD</u>	<u>AALD</u>
Subtests	(<u>n</u> = 1,543)	(<u>n</u> = 338)	(<u>n</u> = 207)	(<u>n</u> = 172)
<u>Factor 1 Verbal Comprehension Path Coefficient</u>				
Information	.80	.75	.81	.35
Similarities	.79	.77	.85	.71
Vocabulary	.82	.83	.89	.74
Comprehension	.69	.75	.84	.74
<u>Factor 2 Perceptual Organization Path Coefficient</u>				
Picture Completion	.60	.68	.80	.55
Picture Arrangement	.50	.57	.77	.51
Block Design	.78	.79	.78	.78
Object Assembly	.69	.67	.78	.79
<u>Factor 3 Freedom From Distractibility Path Coefficient</u>				
Arithmetic	.83	.71	.82	.95
Digit Span	.52	.59	.71	.34
<u>Factor 4 Processing Speed Path Coefficient</u>				
Coding	.61	.63	.71	.51
Symbol Search	.91	.80	.90	.71
<u>Factor Covariances</u>				
F ₁ , F ₂	.69	.75	.82	.33
F ₂ , F ₃	.68	.75	.92	.37
F ₃ , F ₄	.51	.73	.81	.34
F ₁ , F ₃	.75	.86	.88	.59

Table 25 (continued)

F ₂ , F ₄	.59	.64	.79	.30
F ₁ , F ₄	.43	.49	.66	.22

Note. Structure coefficients are reported for each sample in each factor for the four-factor model as well as factor covariances for each sample in the four-factor model. CN = WISC-III Caucasian Sample Without Disabilities; AAN = WISC-III African American Sample Without Disabilities; CD = WISC- III Caucasian Validity Sample With Disabilities; AALD = African American Sample With Learning Disabilities.

Table 26

Model Evaluation Statistics for the Partially Restricted Factor Analysis for the Four-Factor Model

Samples	Goodness-of-fit statistics				
	χ^2	df	Average R	TLI	CFI
WISC-III Caucasian subset ($n = 1,543$)	212.33	48	.03	.97	.98
WISC-III African American subset ($n = 338$)	79.16	48	.03	.97	.98
WISC-III Caucasian subset with disabilities ($n = 207$)	80.83	48	.03	.97	.98
African American Sample with disabilities ($n = 172$)	73.84	48	.06	.92	.94

Note. TLI = Tucker-Lewis index; CFI = comparative fit index; Average R = Average off-diagonal absolute standardized residual.

Appendix F. When the factor structure is applied back upon itself for the CN sample an adequate fit is reported.

The path diagrams for the four-factor model for each sample including factor loadings, factor covariances, error variances of the subtests, and selected fit statistics are reported in Figures 16-19. The four first-order constructs, (Verbal Comprehension, Perceptual Organization, Freedom From Distractibility, and Processing Speed), which are proposed by Wechsler (1991) to represent the theoretical structure of the WISC-III are presented in the ovals to the left of every diagram. Their factor covariances are reported beside the bidirectional arrows connecting pairs of factors. Each grouping of rectangular boxes on the right side of the diagram is the recipient of arrows from a single first-order factor. The rectangles specify the subtests determined by that factor in the four-factor model. The factor loadings of the factor upon each subtest are presented to the left of the rectangular box reporting the subtest's name. The error variance associated with that subtest in the four-factor model is reported to the right of each rectangular box. The chi square, Tucker-Lewis index, and comparative fit index for the partially restricted factor analysis are reported in the small box below the figure.

A comparison of the TLI and CFI obtained for the four-factor model to those obtained for the three and two-factor models reveals trivial differences. Factor loadings from this analysis were examined. They were compared to those reported for the subtests from the partially restricted factor analysis reported for the two-factor model and three-factor model in Figures 3-6 and Figures 10-13. Only trivial differences were noted in all loadings except for Coding, Arithmetic and Information.

Figure 16

The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the Caucasian Sample without Disabilities in the Factor Structure of the Four-Factor Model. ($n = 1,543$)

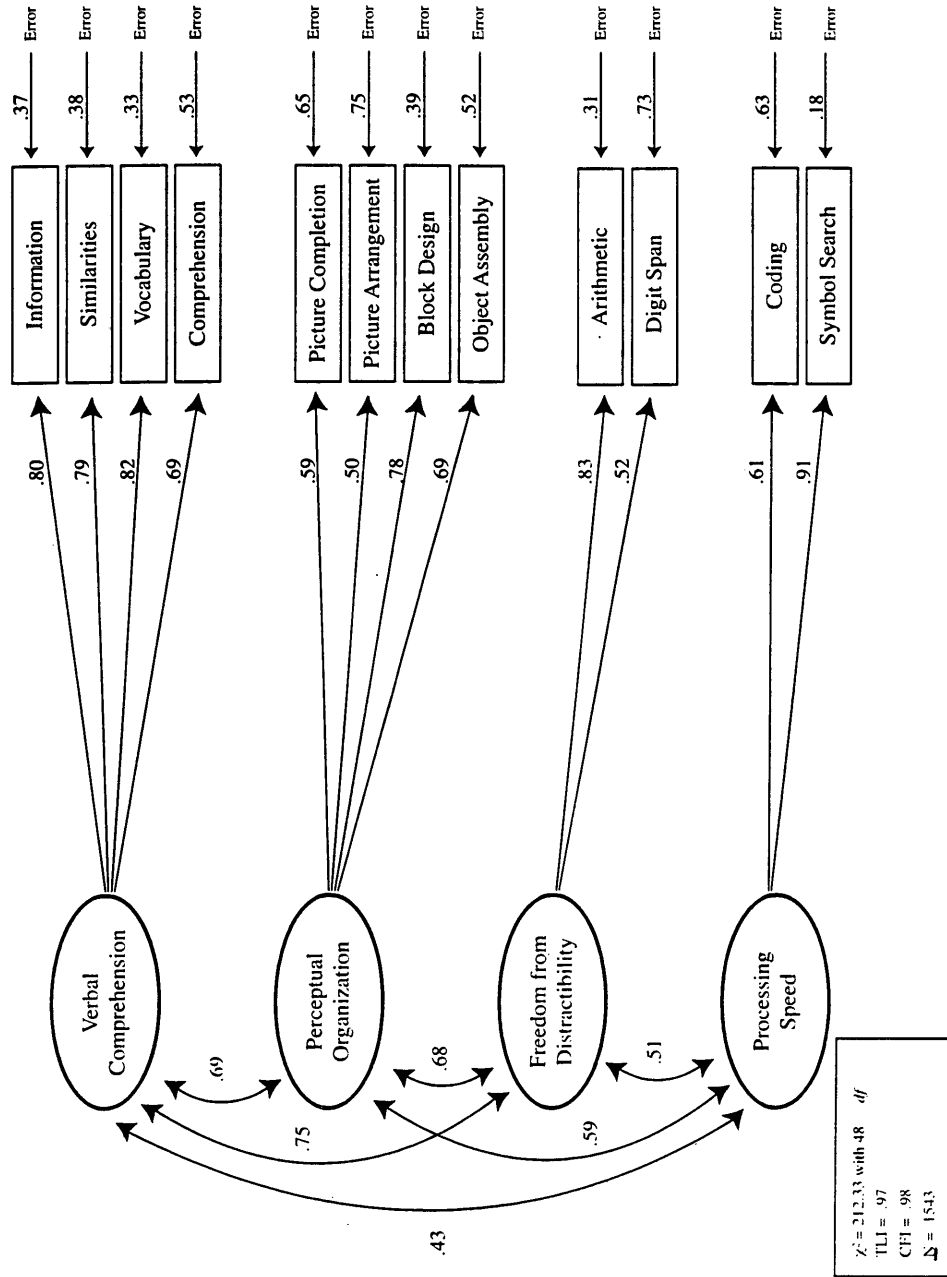


Figure 17

The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample without Disabilities in the Factor Structure of the Four-Factor Model. ($n = 338$)

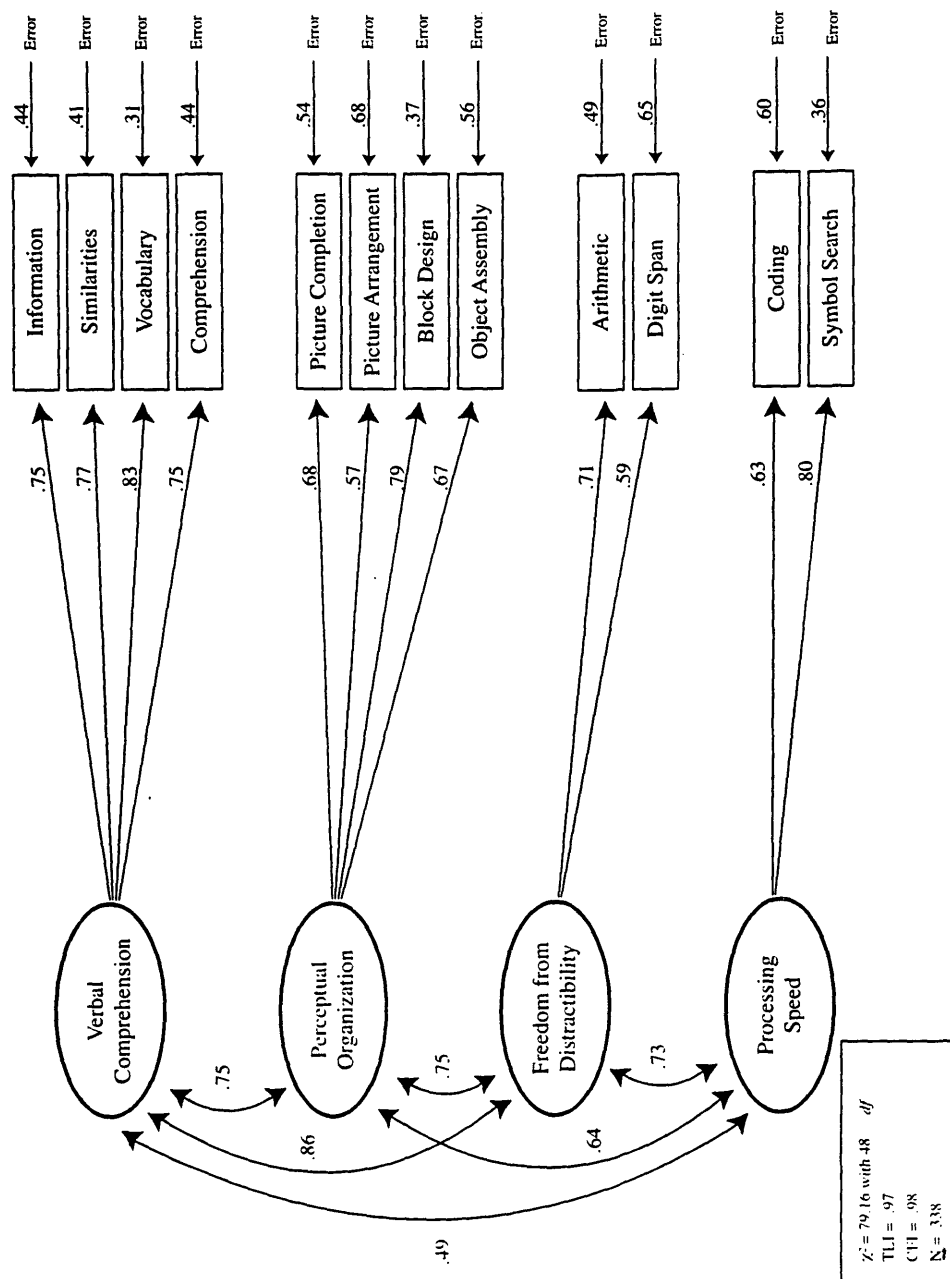


Figure 18

The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the Caucasian Sample with Disabilities in the Factor Structure of the Four-Factor Model. (n = 207)

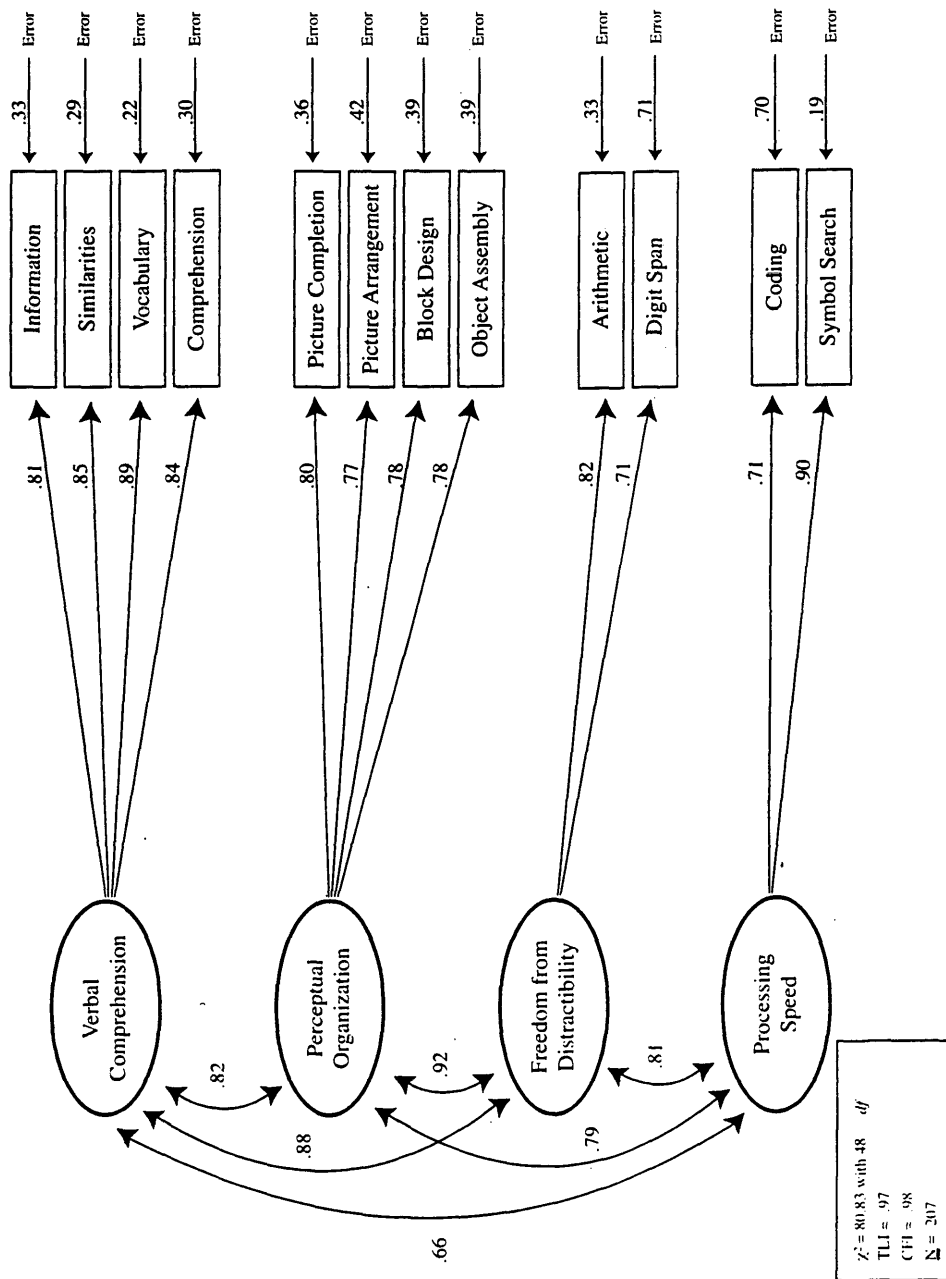
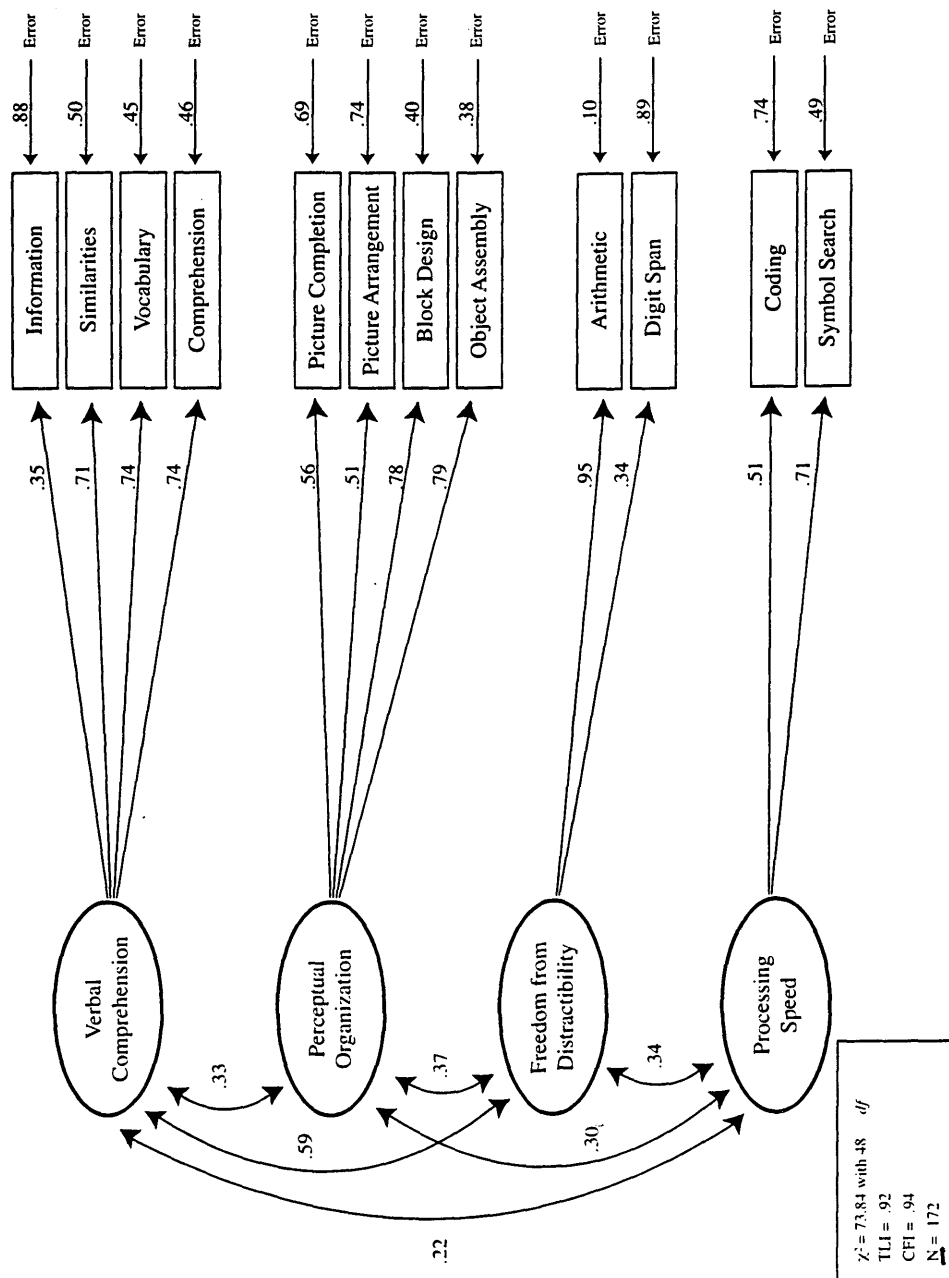


Figure 19

The Path Diagram and Selected Fit Statistics for the Partially Restricted Factor Analysis of the African American Sample with Learning Disabilities in the Factor Structure of the Four-Factor Model. (n = 172)



The loading for Coding on the Processing Speed factor in the four-factor model (.61) was much higher than the loading of Coding on the Freedom from Distractibility factor (.36) in the three-factor model (reported in Table 20) and the loading of Coding on the Performance factor (.36) in the two-factor model (reported in Table 11). The factor loading for Arithmetic in the four-factor model (.83) was slightly larger than the factor loading in the three-factor model (i.e., .80) and higher than the loading in the two-factor model (i.e., .65). These differences would be congruent with statistical procedures used in specifying an additional factor in the initial factor solution for the four-factor model and specifying that Coding and Arithmetic be reconfigured into a factor that was different from the factor specified in the two and three-factor solutions.

Cross Validation of the Factor Structure of the Caucasian Sample (CN) of the WISC-III Standardization Sample to the African American Sample (AAN) of the WISC-III Standardization Sample, the Clinical Validity Sample (CD) of the WISC-III Standardization Sample, and the Sample of African American Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Four-Factor Model

The structure coefficients and factor covariances obtained from the partially restricted factor analysis of the CN sample of 1,543 children were imposed on the covariance matrices of the AAN sample of 338 children, the CD sample of 207 children, and the AALD sample of 172 children, in a fully restricted factor analysis. Fit statistics were obtained and the Lagrange Multiplier Test (LM) was run. Targeted fit indices are reported in Table 27.

Table 27

Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Four-Factor Model

Goodness of Fit Statistics	AAN (<u>n</u> = 338)	CD (<u>n</u> = 207)	AALD (<u>n</u> = 172)
χ^2	113.37	181.62	106.11
<u>df</u>	62	62	62
SRMSR	.07	.29	.20
Average R	.06	.29	.17
TLI	.97	.92	.90
CFI	.97	.93	.90

Note. AAN = WISC-III African American Sample Without Disabilities;

CD = WISC- III Caucasian Validity Sample With Disabilities; AALD = African American Sample With Learning Disabilities; SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

Once again the chi-square value (113.37 with df = 62), the TLI (.97), the CFI (.97), the SRMSR (.07) and the Average R (.06) for the factor structure and loadings of the CN sample imposed on the AAN sample suggest an adequate fit. They offer evidence of the generalizability of the factor structure of the WISC-III four-factor model across ethnicity for samples without disabilities. Although the residuals do reflect slightly less misspecification for the four-factor model than for the three-factor model, they did not reflect improvement when compared to the two-factor model. It

appears there is no advantage statistically to adopting the more complex four-factor model over the two-factor model.

The TLI (.92 for the CD sample and .90 for the AALD sample) and the CFI (.93 for the CD sample and .90 for the AALD sample) reflect some trivial increases over those reported for the three-factor and two-factor models. These may reflect only sampling fluctuations. They suggest marginally adequate fit again. Examination of the residuals as reflected in the SRMSR (.30 for the CD sample and .17 for the AALD sample) suggests the addition of the fourth factor did not improve the reproduction of the correlations over the three-factor or two-factor model. The two-factor model reproduced the correlations most accurately across all samples of children with disabilities. The issue of differences in variances for subtest scores was noted again regarding the CD sample and the AALD sample in examining the distribution of standardized residuals. No improvements in malfitting parameters could be made using the results from the LM test without seriously violating the assumptions of the four-factor model.

Cross-Validation of the Factor Structure of the African American Sample (AAN) of the WISC-III Standardization Sample to the African American Sample of Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the Four-Factor Model

The factor loadings and factor covariances obtained from the partially restricted factor analysis of the AAN sample of 338 children were imposed on the covariance matrix of the AALD validity sample of 172 children in a fully restricted factor analysis. Fit statistics were obtained and results related to fit statistics are

presented in Table 28. A comprehensive report of fit statistics appears in Table 2 in Appendix F. The Lagrange Multiplier Test (LM) was also conducted.

Table 28

Model Evaluation Statistics for the Factor Structure and Loadings of the African American Sample without Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Four-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AALD</u>
χ^2	123.52
<u>df</u>	62
SRMSR	.21
Average R	.18
TLI	.86
CFI	.86

Note. AALD = African American Sample with Learning Disabilities;

SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

Goodness-of-Fit Statistics

As reported in Table 28, the TLI (.86) and CFI (.86) reflect a marginal fit to the data. The marginality of fit suggests somewhat greater misfit in the model of the AAN loadings to the AALD sample for the four-factor model than the three-factor model and the two-factor model. The other goodness of fit indices reflect this consistently as well.

Interpretation of the slight decrease in fit across models is complicated by the fact that sample size of the AALD group ($n = 172$) was greatly diminished for the four-factor covariance matrix in comparison to the covariance matrix for the two-factor and three-factor models. Model misspecification is consistent with that reported for the two-factor and three-factor models. Differences in variance for some of the subtests probably accounted for much of the error.

The inspection of the cumulative multivariate LM test statistics along with their accompanying univariate increments revealed four less substantial parameters that would involve adding paths from observed variables to another factor, resulting in serious violations of the model. Individually, none of them improved the chi-square substantially, suggesting the model fit was not very good.

Cross-Validation of the Factor Structure of the Caucasian Validity Sample (CD) of the WISC-III Standardization Sample to the African American Sample of Children with Learning Disabilities (AALD) in a Fully Restricted Factor Analysis for the

Four-Factor Model

The factor loadings and factor covariances obtained from the partially restricted factor analysis of the CD sample of 207 children were imposed on the covariance matrix of the AALD validity sample of 172 children in a fully restricted factor analysis. Fit statistics were obtained and results related to fit statistics are presented in Table 29. A comprehensive report of all fit statistics appears in Table 3 in Appendix F. The Lagrange Multiplier Test (LM) was also conducted.

Table 29

Model Evaluation Statistics for the Factor Structure and Loadings of the Caucasian Sample with Disabilities Imposed on the African American Sample with Learning Disabilities in a Fully Restricted Factor Analysis for the Four-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AALD</u>
χ^2	174.88
<u>df</u>	62
SRMSR	.62
Average R	.59
TLI	.73
CFI	.75

Note. AALD = African American Sample with Learning Disabilities; SRMSR = standardized root mean square residual; Average R = average off-diagonal absolute standardized residual; TLI = Tucker-Lewis index; CFI = comparative fit index.

Goodness-of-Fit Statistics

As reported in Table 29, the TLI (.73) and CFI (.75) reflected a poor fit to the data, consistent with that reported for the two-factor model before respecification and the three-factor model. The consistently poor statistics obtained with respect to the SRMSR (.62) and the Average R (.59) suggest considerable misspecification in the model of the CD loadings to the AALD sample. Because the sample size of the AALD sample differed for the four-factor model, some comparisons of statistics for the two-factor model and three-factor model could not be made. Differences in variances for the subtests were considered to be a major source of error.

LM Test

The inspection of the cumulative multivariate LM test statistics along with their accompanying univariate increments revealed four less sizable misfitting parameters, all involving adding a path from one of the observed variables to another factor, resulting in a serious violation of the model. None of the individual additions would substantially improve the chi-square statistic.

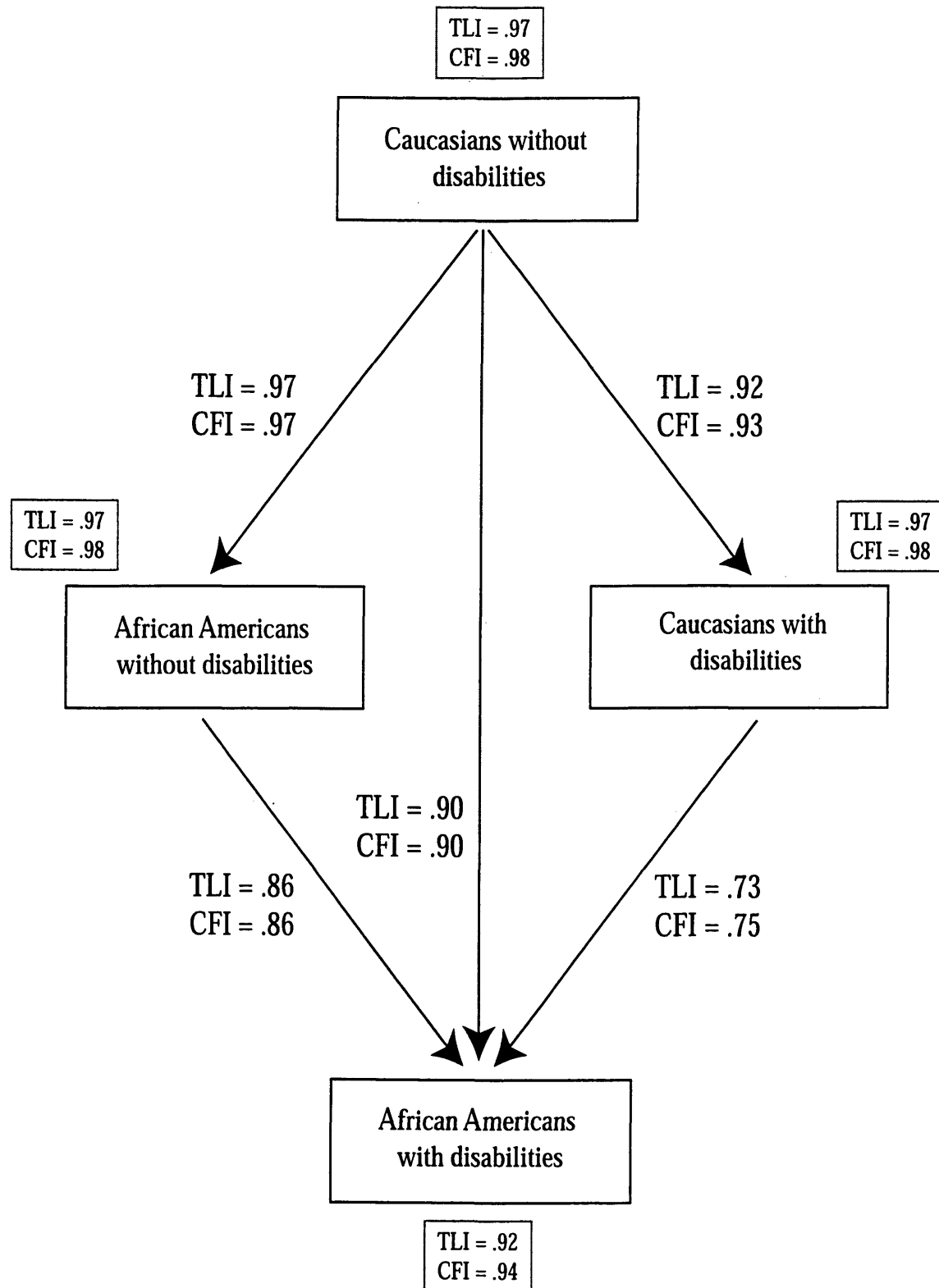
Overview of Results for the Four-Factor Model

Assessment of Overall Model Fit

Figure 20 entitled “Overview Model Fit of Four-Factor Model Across Samples” illustrates the overall model fit of the four-factor model that Wechsler (1991) proposed to represent theoretical structure of the WISC-III. The critical nature of the imposition of the factor loadings from the Caucasian samples initially upon all other samples is emphasized by its position at the apex of the figure as its application upon the three samples represented below it. The imposition of the factor loadings from one sample upon another in a fully restricted factor analysis in every analysis is indicated by a unidirectional arrow. The tail of the arrow emanates from the sample where factor loadings are being imposed upon another sample. The head of the arrow rests against the oval representing the sample upon which the factor loadings have been imposed. The TLI and CFI values presented in proximity to the arrow heads represent the fit statistics in the fully restricted factor analysis for the four-factor model. The TLI and CFI scores that are enclosed in boxes reflect values for the partially restricted factor analyses performed previously upon each sample.

Figure 20

Overall Model fit of Four-Factor Model Across Samples.



These analyses had generated the factor loadings imposed in the fully restricted factor analyses illustrated in this figure. The broad conclusion seems to be, so far, that the CN results generalize best across all groups, for all models.

Analysis of the Model Evaluation Statistics Across Samples

The model evaluation statistics for the partially restricted factor analysis of the four samples (consistently including the data from the four-factor model AALD sample) across the two-factor, three-factor and four-factor models are represented in Table 30 in order to determine which model provides the best fit. Use of the data from the four-factor model AALD sample made consistent comparison possible for each sample in each model. Criteria for comparing models across samples in this study were guided by the standards articulated by MacCallum et al. (1992) as well as those of Thorndike (1978) reported previously.

A desirable outcome in CSM [covariance structure modeling] analysis is to find that the model under investigation fits well, that it cannot be simplified substantially without significant loss of overall fit and that its fit cannot be improved to any great extent by making the model more complex (MacCallum, 1992, p. 490).

The model evaluation statistics for the four samples across the three models were examined. Results suggest the inclusion of each additional factor improved the goodness-of-fit for the CN and CD samples consistently but modestly as reflected in the ratio of the chi-square divided by its degrees of freedom and the SRMSR. The inclusion of the fourth factor resulted in slightly greater improvement in fit than the two or three-factor model for the CN and CD samples. Residuals accordingly were trivially lower as a result of the addition of each succeeding factor, with the four-factor model revealing the smallest misspecification for the CN and CD samples.

Table 30**Model Evaluation Statistics Across Samples and Models for the Partially Restricted Factor Analyses**

<u>Model Type</u>	<u>Goodness-of-fit statistics^a</u>				<u>Model improvement statistics^a</u>	
	χ^2	df	χ^2/df	SRMSR	TLI	CFI
CN ($n = 1,543$) ^b						
Two-Factor	252.73	34	7.43	.04	.95	.96
Three-Factor	227.05	41	5.54	.04	.96	.97
Four-Factor	212.33	48	4.42	.03	.97	.98
AAN ($n = 338$) ^b						
Two-Factor	71.61	34	2.11	.04	.96	.97
Three-Factor	83.67	41	2.04	.04	.96	.97
Four-Factor	79.16	48	1.65	.04	.97	.98
CD ($n = 207$) ^b						
Two-Factor	84.02	34	2.47	.04	.95	.96
Three-Factor	78.75	41	1.92	.04	.95	.97
Four-Factor	80.83	48	1.68	.03	.95	.97
AALD ($n = 172$) ^b						
Two-Factor	56.15	34	1.65	.07	.93	.95
Three-Factor	68.57	41	1.67	.08	.91	.94
Four-Factor	73.84	48	1.54	.07	.92	.94

Note. The results for all models are shown for comparison purposes. They are the same as reported in Figures 3-6, 10-13, and 16-19. ^aSRMSR = standardized root mean square residual; TLI = Tucker-Lewis index; CFI = comparative fit index. ^bCN = WISC-III Caucasian Sample Without Disabilities; AAN = WISC-III African American Sample Without Disabilities; CD = WISC- III Caucasian Validity Sample With Disabilities; AALD = African American Sample With Learning Disabilities.

The AAN and AALD samples did not show consistent improvement in goodness-of-fit statistics as additional factors were added. Trivial deterioration in fit was noted when the statistics for the three-factor model were compared to those generated for the two-factor model. However, fit for the four-factor model was slightly better than the three-factor model and the two-factor model as reflected in the chi-square/df ratio and SRMSR.

Model improvement statistics reflect a somewhat more diverse pattern with regard to fit. The statistics for the CN and CD samples improved slightly but steadily with the addition of each successively more complex factor structure. The one exception is the TLI for the CD sample in the three-factor model. The AAN sample, by contrast, did not reflect model improvement from the two-factor to three-factor model but did reflect trivial model improvement from the two-factor model to the four-factor model. Overall, the four-factor solution marginally provided the best fit across these three samples. The model improvement statistics reflected only modest gains as factor complexity increased to four factors.

The picture is more complicated in regard to the AALD sample. The model improvement statistics did not reflect model improvement from the two-factor model to the three-factor model. The TLI for the four-factor model was only trivially lower than for the two-factor model.

Comparability of Factor Loadings Across Other Independent Samples

Samples of African American Children with Disabilities

Factor loadings in the three-factor solution of the Kush et al. (in press) samples, the AALD, and CD samples are reported together in Table 31. Model evaluation statistics chosen to be emphasized and reported in a four-factor solution for the Kush et al., for the AALD and CD samples are reported in Table 32. Three common statistics reported in comparing samples from the two studies were included. Because the chi-square is unduly influenced by sample size, and there is variation in the sizes of these three samples, the CFI and SRMSR were considered to be more meaningful model evaluation statistics to be inspected. In making comparisons, the issue of composition of the samples must be considered additionally. The AALD sample reflects children designated as having learning disabilities. The CD and Kush et al. (in press) samples are mixed. The latter, in fact, includes 81 referred subjects who were found to be ineligible for special education services.

Samples of Other Ethnically Diverse Samples with Disabilities

Factor loadings in the four-factor solution for the Konold, Kush, and Canivez sample (1997), the AALD and the CD samples are reported in Table 33. Factor loadings for subtests from the two most robust factors (i.e., Verbal Comprehension and Perceptual Organization) in the four-factor solution representing Wechsler's (1991) theoretical structure from the confirmatory factor analysis of the Kush et al. (in press) sample are reported. Also, factor loadings for comparable subtests in the

Verbal Comprehension and Perceptual Organization factor in the four-factor solution representing Wechsler's (1991) theoretical structure from the partially restricted factor analysis of the AALD and CD samples are reported in Table 34. Both the CD and Kush et al. (in press) samples are mixed clinical samples, whereas the AALD sample includes only children designated as having learning disabilities. None of the samples are stratified. The entire composition of the sample is not consistent. Sample sizes vary. The Kush et al. (in press) sample being considerably larger than the CD or AALD sample. All of these issues must be reviewed in examining factor loadings.

Table 31

Factor Loadings for the Subtests Composing the Verbal Comprehension and Perceptual Organization Factors for the African American Sample with Learning Disabilities, Caucasian Validity Sample with Disabilities and the Kush et al., (in press), Samples in Three-Factor Solutions

Subtest	AALD	CD	Kush et al.
Verbal Comprehension Factor			
Information	.54	.81	.62
Similarities	.66	.85	.62
Vocabulary	.77	.88	.84
Comprehension	.69	.84	.68
Perceptual Organization Factor			
Picture Completion	.64	.81	.55
Picture Arrangement	.42	.76	.44
Block Design	.55	.78	.77
Object Assembly	.65	.78	.66

Note. CD = WISC- III Caucasian Validity With Disabilities; AALD = African

American Sample with Learning Disabilities.

Table 32

Model Evaluation Statistics for the Partially Restricted Factor Analysis of the African American Sample with Learning Disabilities, and the Caucasian Validity Sample with Disabilities, and the Confirmatory Factor Analysis of the Kush et al., (in press), Samples in the Four-Factor Model

Goodness of Fit Statistics	AALD	CD	Kush et al.
χ^2	73.84	80.83	151.50
<u>df</u>	48	48	48
CFI	.94	.97	.93
SRMSR	.07	.03	.05

Note. CD = WISC- III Caucasian Validity With Disabilities; AALD = African American Sample with Learning Disabilities; CFI = comparative fit index, SRMSR = standardized root mean square residual.

Table 33

Factor Loadings for the Four-Factor Model for the African American Sample with Learning Disabilities, and the Caucasian Validity Sample with Disabilities, and Konold, Kush, and Canivez (1997) Samples

Subtest	<u>Konold, et, al. Samples</u>				
	AALD	CD	Mixed Sample 1	Mixed Sample 2	Mixed Sample 3
Verbal Comprehension Factor					
Information	.35	.81	.80	.71	.75
Vocabulary	.71	.85	.86	.83	.79
Similarities	.74	.89	.79	.81	.81
Comprehension	.74	.84	.79	.70	.68
Freedom From Distractibility Factor					
Arithmetic	.95	.82	.84	.68	.80
Digit Span	.34	.71	.64	.38	.58
Perceptual Organization Factor					
Block Design	.56	.80	.82	.72	.65
Object Assembly	.51	.77	.71	.61	.67
Picture Completion	.78	.78	.74	.61	.62
Picture Arrangement	.79	.78	.69	.65	.59
Processing Speed Factor					
Coding	.51	.71	.59	.43	.46
Symbol Search	.71	.90	.79	.99	.99

Note. CD = WISC- III Caucasian Validity With Disabilities; AALD = African

American Sample with Learning Disabilities.

Table 34

Factor Loadings for Selected Subtests in the Four-Factor Model for the African American Sample with Learning Disabilities, the Caucasian Validity Sample with Disabilities, and the Confirmatory Factor Analysis of the Kush et al., (in press), Samples

Subtest	AALD	CD	Kush
Verbal Comprehension Factor			
Information	.54	.81	.71
Similarities	.66	.85	.75
Vocabulary	.77	.88	.81
Comprehension	.69	.84	.75
Perceptual Organization			
Block Assignment	.55	.78	.81
Picture Completion	.64	.81	.49
Picture Arrangement	.42	.76	.54
Object Assembly	.65	.78	.69

Note. CD = WISC- III Caucasian Validity Sample With Disabilities;

AALD = African American Sample with Learning Disabilities

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Overview

The WISC-III is the most widely used test of general intelligence in the assessment process employed to determine whether school-age children have a learning disability. Numerous studies have investigated the factor structure of the WISC-III standardization sample. The WISC-III manual suggests a four-factor model represents the data from the WISC-III standardization sample most accurately. Additionally, it recommends Index scores from these four factors be used in interpretation of test results. However, other researchers have identified several other models, including a one-factor solution, a two-factor solution, a three-factor solution, and an alternative four-factor solution.

The ethnic group experiencing the most rapid growth with respect to identification of learning disabilities is African American school-age children. The reliance of school psychologists on the WISC-III in the assessment of ethnic minority children suspected of having a learning disability has prompted test users to question whether the factor structure of the WISC-III standardization sample generalizes to these children. A basic source of confusion in addressing this question arises from the failure of the authors of the WISC-III to publish information about the generalizability of the factor structure of the complete standardization sample across

its different subsamples. These include its ethnic subsamples without disabilities as well as the subsample which consists of both children “at risk” and children in mainstream programs with disabilities.

Studies involving independent samples of African American children with disabilities, including children with learning disabilities have yielded results somewhat difficult to interpret. Efforts to disaggregate the effects of disability status from ethnicity have appeared to be incomplete in these studies. Additionally, the use of mixed samples with disabilities has contributed a methodological problem in some studies.

The present study attempted to evaluate the effects of both ethnicity and disability status. This was begun initially by investigating the question of the generalizability of the factor structure of the WISC-III in the three most commonly used models (i.e., two-factor, three-factor, and four-factor) across the Caucasian and African American samples of the WISC-III standardization sample without disabilities. Subsequently, the generalizability of the factor structure of the samples without disabilities was investigated in relation to two samples with disabilities, the WISC-III Caucasian mixed validity sample and an independent sample of school-age African American children designated with learning disabilities (the latter having been included after it was determined the factor structure was consistent across all age levels of the sample).

A second fundamental question addressed in this research was which independent variable, ethnicity or disability status, caused the greater stress in model applications. Examination of goodness- of-fit indices, especially the SRMSR and

Average R that indicate degree of misspecification, provided evidence of which independent variable causes greater stress. Lastly, the question of which of the three models provided the best fit across the two samples without disabilities and the two samples with disabilities was investigated by examining model evaluation statistics from the partially restricted factor analyses of the four samples over the three models used.

The following discussion begins with a summarization of the major findings based on the statistical analyses that were conducted. The statistical evidence regarding the generalizability of the factor structure for the two-factor, three-factor, and four-factor models across the four samples will be reviewed initially. Subsequently, the model evaluation statistics, comparing the four samples will be reviewed. The results of these analyses will be related to the results of other studies involving samples with disabilities. Additionally, the results will be related to existing literature on current assessment practices. Limitations of the current study and recommendations for future research will conclude the chapter.

Summary of Statistical Analyses

Overview

The methodology used to investigate the generalizability of the factor structure of the WISC-III from the WISC-III standardization sample to the African American sample with learning disabilities involved a technique call cross-validation of covariance structure models (CVCSM), in an extension of the work of Allen and Thorndike (1995a, 1995b). Covariance matrices were generated for each of the four

samples investigated using their matrices of subtest intercorrelations and standard deviations. Using CVCSM, evidence of the generalizability of the factor structure was obtained from the examination of model evaluation statistics reflecting the goodness-of-fit of the fully restricted factor analytic models generated from the partially restricted factor analyses and the LM test.

If the fit did not degrade appreciably when applied in a fully restricted factor analysis from the Caucasian sample without disabilities to the African American sample without disabilities and the African American sample with disabilities, evidence was presented that the factor structure was reliable across ethnicity. If the fit did not degrade appreciably when applied in the fully restricted factor analysis from the samples without disabilities to the samples with disabilities statistical evidence was presented that the factor structure was reliable across disability status for that model.

Lastly, the question of which model provided the best fit across the four samples was addressed by examining the model evaluation statistics for each sample across each model in the partially restricted factor analyses, comparing increments of fit. The statistics revealing the least misspecification as indicated by the SRMSR while at the same time reflecting the greatest reduction in chi square and the chi square/degrees of freedom ratio, as well as robust fit in the TLI and CFI, were considered to represent the most generalizable factor structure model.

Results for the Two-Factor Model

The factor structure of the two-factor model for the WISC-III Caucasian sample without disabilities (CN) replicated over the WISC-III African American

sample without disabilities (AAN), provided an excellent fit for the model, and provided evidence to support the factor structure as being group-invariant across ethnicity for the samples without disabilities. When the factor structure for the two samples without disabilities was imposed upon the two samples with disabilities acceptable fit statistics were obtained, although greater degradation in fit statistics was noted. The greatest degradation in fit was reported, however, when the factor structure for the CD sample was imposed on the AALD sample.

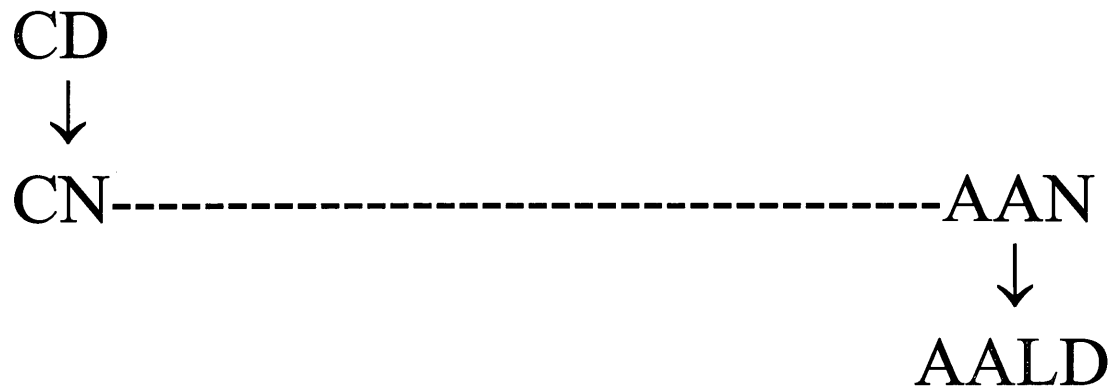
When the results of the LM test were reviewed it was noted in every case that the major malfitting parameter specified for both samples with disabilities involved freeing the factor covariance. The consistent finding is related to the differences in the magnitude of the correlations in the two samples. There is a change in the signs of the correlations in different directions for the two disability groups in comparison to the magnitude of the correlations for the CN and AAN samples, as illustrated in Figure 21.

The AALD sample has overall smaller standard deviations than the CN and AAN samples. The CD sample has considerably larger standard deviations than the CN and AAN samples. Larger variances serve to elevate the correlations and covariances while smaller variances severely depress correlations and covariances. When the factor covariance is freed and the model evaluation statistics are rerun, in every case they reflect acceptable fit for the two-factor model.

The work of MacCallum, Roznowski, and Necowitz (1992), in which they demonstrated the difficulty in replicating models over different samples, provided a

Figure 21

Differences in Average Magnitude of the Correlation Matrices of the
Sample with Disabilities in Comparison to the Samples without Disabilities.



rationale for concluding the factor structure of the two-factor model did appear to be group invariant across ethnicity. Additionally, it provided a rationale for concluding the factor structure did appear to be invariant across disability status for the two-factor model. When differences in factor covariance were taken into consideration, excellent fit was achieved for the two-factor model for both samples with disabilities in the two-factor model.

Results for the Three-Factor Model

The factor structure of the three-factor model for the WISC-III Caucasian sample without disabilities (CN) replicated over the WISC-III African American sample without disabilities, providing an adequate fit for the model. The fit suggested the construct validity of the WISC-III for the three-factor model was group-invariant across ethnicity in these two samples of children without disabilities. However, the increase in the size of the residuals from the two-factor model to the three-factor model presented evidence the addition of the third factor did not reproduce the covariances more accurately.

The factor structure of the three-factor model for the WISC-III CN sample when imposed on the CD sample provided evidence of factor structure reliability for that model. Trivially higher TLI and CFI values were produced for the three-factor model than the two-factor model, suggesting the addition of one additional factor did not substantially improve fit. Additionally, the increase in the size of the residuals suggested the added third factor did not reproduce the covariances more accurately than the two-factor model. Also, the results of the LM test did not offer any suggestions that would result in model improvement without seriously violating

assumptions about the model, suggesting the misspecification was generalized across the model.

When the factor structure of the three-factor model for the WISC-III CN sample and the AAN sample were imposed on the AALD sample, the generalizability of its factor structure was affirmed as reflected in fit statistics. The TLI and CFI values were trivially lower than those reported for the two-factor model. Misspecification as reflected in examination of residuals was trivially larger than it was for the two-factor model. The results of the LM test did not offer any suggestions that would result in model improvement, even taking into account suggestions that seriously violated assumptions about the model. The increase in model complexity did not improve fit appreciably.

The imposition of the factor structure of the CD sample on the AALD sample for the three-factor model was problematic. It resulted in greater misspecification than it did for the two-factor model. Additionally, the LM test did not provide suggestions for respecification that would substantially improve the model fit, as had been the case for the two-factor model. The three-factor model did not generalize from the CD sample to the AALD sample. The significant loss of overall fit in this more complex model for these samples with disabilities suggests the model does not provide a plausible explanation.

No apparent advantage was observed in imposing the three-factor model on any of the samples when compared to results for the two-factor model. Although the factor structure replicated across ethnicity in both samples without disabilities, the residuals did not show improvement over the two-factor model for the AAN sample.

Statistics involving both samples with disabilities suggested the three-factor model provided a marginally poorer fit than the two-factor model. This may be the result in part of sampling fluctuation. The AALD sample in the three-factor model was somewhat smaller than the sample used in the two-factor model. Additionally, however, poorer fit may be the result in part of model complexity. MacCallum et. al., (1994) suggests increasing model complexity makes cross validation more difficult to achieve.

Results for the Four-Factor Model

The factor structure of the four-factor model for the WISC-III Caucasian sample without disabilities (CN) replicated over the WISC-III African American sample without disabilities (AAN), providing an excellent fit for the model, as reflected in the TLI and CFI. The fit provided evidence to support the construct validity of the WISC-III in that the four-factor model appeared invariant across ethnicity in these two samples of children without disabilities. However, the size of the residuals for the four-factor model when compared to the two-factor model suggested the addition of the factors in the four-factor model did not improve the accuracy of reproductions of the intercorrelations of the subtest scores. The two-factor model provided a more parsimonious explanation with greater generalizability. There appears to be no advantage statistically to adopting the four-factor model over the two-factor model.

The factor structure of the four-factor model for the WISC-III CN sample when imposed on both samples with disabilities generalized about as well as it did for the three-factor model. It did not provide substantially better model fit than the two-

factor model. Examination of the statistics when the four-factor solution from the AAN sample was imposed on the AALD sample revealed trivial differences when compared to the three-factor model and the two-factor model without respecification. The generalizability of the four-factor structure was marginally affirmed. It was noteworthy that the results of the LM test for the four-factor model offered no suggestions for respecification that would substantially improve the model fit. This suggested the error is not the result of one misfitted parameter, but rather was spread across the entire model. The four-factor model offers no advantages statistically over either of the other models. It does not provide a better explanation of the data.

When the four-factor solution from the CD sample was imposed on the AALD sample, the degree of misspecification was the most severe across the three models investigated. The four-factor model did not appear to generalize across disability status from the CD sample to the AALD sample. The results of the LM test offered no suggestions for respecification that substantially improve the model, suggesting the error is not the result of one misfitted parameter, but rather is spread across the entire model. A single four-factor model, does not generalize across the two samples with disabilities. It does not provide an adequate explanation for the data.

Conclusions

The Question of Generalizability of Factor Structure Across Samples

The goodness of fit results presented in Tables 15, 17, 22, 23, 27, and 28 overall suggested remarkable consistency in the factor structure of the models across ethnicity in samples without disabilities for the two-factor, three-factor, and four-

factor models. The factor structure of the three models was robust and reliable from the CN sample to the AAN sample. The factor structure of the two-factor model generalized from the CN and AAN samples to the AALD sample. The factor structure of the two-factor model generalized from the CN sample to the CD sample.

When the factor structure of the CD sample was imposed upon the factor structure of the AALD sample in the two-factor model, the fit statistics were less robust. When the factor pattern was employed in the respecified two-factor model in which the factor covariances were freely estimated, the fit statistics generated were excellent. They were weaker for the three-factor model and weaker still in the four-factor model than in the two-factor model without respecification. The residuals in the four-factor model were especially large. The factor structure of the CD samples in the three-factor and four-factor solutions did not generalize across the AALD sample acceptably, indicating these solutions did not provide an accurate statistical explanation for the data.

A complication in interpreting results was the fact the AALD samples differed for each of the three models because not all subtests had been administered to every child. Sampling fluctuation may have accounted for some of the error in specification of factor loadings. Restriction in range in all of the subtest scores of the AALD samples was an important consideration. Additionally, the much larger variances of subtests of the CD sample elevated the correlations and covariances, whereas the smaller variances of the AALD sample served to reduce the correlations and covariances considerably.

The Question of which Independent Variable Causes Greater Model Stress

The finding that the generalizability of the factor structure of the three models across disability status was more problematic than across ethnicity prompted further examination of the data. Greater stress in every model application was noted from the samples without disabilities to the samples with disabilities. The greatest stress was noted in the generalizability of the factor structure from the CD to the AALD sample, however. The size of the covariances among factors for the samples emerged as a very important consideration in understanding this finding.

An overview examination of factor covariances for each sample across each model distinguished the two samples with disabilities from the two samples without disabilities with regard to factor covariances. The CD sample generated the highest factor covariances consistently. The factor covariances of the CN and AAN samples were more closely aligned in the middle of the distribution of factor covariances and reflected lower factor covariances consistently than the CD sample. The factor covariances for the AALD sample, by contrast, were consistently far lower than the other three samples across all three-factor structure models.

Closer interpretation of factor covariances within samples yielded an interesting finding of very high factor covariances for some samples of the three and four-factor models. The issue of overfactoring of the WISC-III was raised initially by Kamphaus et al. (1994) in their confirmatory factor analysis of the entire WISC-III standardization sample for the same three factor models investigated in the present study. Kamphaus et al. reported a high degree of factor covariance for the Verbal Comprehension and Freedom from Distractibility factors in the three-factor model for

the WISC-III standardization sample. Additionally, to a slightly lesser extent in the four-factor model they noted the high degree of factor covariance for the Verbal Comprehension and Freedom from Distractibility factors. They suggested these findings may represent a case of overfactoring for some age groups especially.

In the present study a high degree of factor covariation for the Verbal Comprehension factor and the Freedom from Distractibility factor (i.e., .83 in the three-factor model and .86 in the four-factor model) was reported for the AAN sample. Additionally, a high degree of factor covariation was noted for the CD sample across all models. It was especially noteworthy in the factor covariance of the Verbal Comprehension factor with the Freedom from Distractibility factor in the three-factor model (i.e., .86) and four-factor model (i.e., .88) and the factor covariance of the Perceptual Organization factor with the Freedom from Distractibility factor for the three-factor model (i.e., .92) and the four-factor model (i.e., .92). The high degree of factor covariance in these instances suggested the possibility of overfactoring with respect to the three and four-factor models for the AAN and CD samples, offering another claim of evidence that the factor structure of the two-factor model provides the most generalizable factor solution.

The other finding with regard to factor covariances involved results of the LM test in the two-factor solution. For the factor structure of the two-factor model, alone, the LM test consistently suggested that freeing of the parameter involving factor covariance resulted in marked improvement in fit as reflected in respecified fit statistics for the AAN, CD, and AALD samples. The improvement, confirmed in the

results reported in Tables 16, 18, and 19, was noted in every fully restricted factor analysis across samples, both with regard to ethnicity and disability status.

The Question of which Model Provides the Best Fit

Examination of model evaluation statistics revealed trivial differences in increases in the TLI and CFI coupled with trivial decreases in the SRMSR and Average R in the three and four-factor models for the CN sample. The four-factor model reflected trivially higher values of the TLI and CFI of the three models for the CN, CD and AAN samples. In the AALD sample the TLI and CFI were highest for the two-factor model but differences were trivial in model evaluation statistics across the three models. Sample fluctuations alone could account for the trivial differences in model evaluation statistics over the four samples.

The SRMSR and Average R, indicators of “badness of fit”, although lowest in the four-factor model, reflected trivial differences across the three models. One would expect a decrease in misspecification with the addition of more factors. The decreases observed in this study however, must be characterized as being trivial when one takes into account the degree of factorial complexity that has been introduced in the three-factor and four-factor models.

Additionally, in the four-factor model for the AALD sample, examination of the factor loadings for the Freedom from Distractibility factor reveals a loading of .95 for Arithmetic. For this sample, therefore, Arithmetic was capturing the latent factor in totality, raising the question of overfactoring.

The conclusion, reached from examination of fit statistics and model evaluation statistics, was that the two-factor model did provide the most adequate fit.

The fundamental fit of the two-factor model in and of itself was substantial. Little room for improvement existed to begin with in fit statistics based upon the model fit of the two-factor model. The trivial differences in increases in model improvement statistics (that may be due solely to sampling fluctuations) and fit statistics (that may be the result of the introduction of additional factors) do not warrant increasing model complexity.

Relating Findings of the Present Study to Other African American Samples with
Disabilities

Questions always arise about comparability of results from independent samples in data analysis. One way of evaluating comparability of results is by comparing factor loadings obtained from the same subtest scores across different samples.

The subtest scores from the Kush et al., (in press) mixed, referred sample involving 348 African American children were subjected to both exploratory and confirmatory factor analyses. Factor loadings were reported only for the exploratory factor analysis. It supported a three-factor model. This model departed from the Kaufman three-factor model in that the third factor emerging was identical to the Processing Speed factor which was identified in the analysis of the WISC-III standardization sample by its test developers (Wechsler, 1991). Freedom from Distractibility did not emerge as a factor in the exploratory factor analysis of the Kush et al. sample. Factor loadings from the Verbal Comprehension and Perceptual Organization factors in the Kush et al., sample were compared initially to the factor loadings obtained from the CD and AALD samples for the subtests composing the

Verbal Comprehension and Perceptual Organization factors in the three-factor model. The factor loadings for the subtests composing the Verbal Comprehension factor and Perceptual Organization factor in the three-factor models generated for the AALD, CD and Kush et al., samples were reported in Table 31 in Chapter IV.

Inspection of the factor loadings from the Kush et al. (in press), sample suggests they are remarkably congruent with factor loadings for the AALD sample. Both samples of African American children with disabilities differ about as much from the CD sample with respect to these factor loadings, and in both cases the loadings are lower. The consistency of these loadings provides evidence of the replicability of the findings from the current study with respect to ethnicity and disability status. One reason the correlations might be lower is that there is less influence of a general factor. Another might simply be how the samples were selected.

The CFI and SRMSR from the confirmatory factor analysis of the Kush et al. (in press) sample, in the four-factor model, the CFI and SRMSR, were compared to those generated for the AALD sample and CD samples in the four-factor model (refer to Table 32 in Chapter IV). Kush et al. identified this model as providing the best fit of all models investigated in their confirmatory factor analysis.

The trivial differences noted across these statistics probably are attributable in part to sampling fluctuations as well as the heterogeneous disability status of the Kush et al., and CD groups. Although Kush et al. concluded the four-factor structure provides the best fit in the confirmatory factor analysis, their general conclusions are congruent with those of the current study. They offer three types of evidence to

bolster this conclusion. First, there were differences in the results of the exploratory and confirmatory factor analyses, yielding inconsistent evidence regarding factor structure. Second, the small size of the Freedom from Distractibility and Processing Speed factors in the four-factor model suggested they are accounting for very little variance. Lastly, lack of sufficient salient variables per factor for these two factors raised questions about the validity of the four-factor model. “Interpretation beyond global, verbal, and performance dimensions should be undertaken with caution” (Kush et al., p. 15).

The Kush and Watkins (1997) mixed sample of 161 student African American children with disabilities was subjected to a maximum likelihood/direct oblique exploratory factor analysis. Although the methodologies differ between that sample and the AALD sample in the present study that involved a maximum likelihood/oblique partially restricted confirmatory factor analysis, the factor loadings for the two samples were compared.

Factor loadings showed remarkable congruence over the two samples with four exceptions. The loading for Information was .59 for the AALD sample and .87 for the Kush and Watkins (1997) sample. The loading for Block Design was .59 for the AALD sample and .92 for the Kush and Watkins sample. The loading for Comprehension was .62 for the AALD sample and .78 in the Kush and Watkins sample. Additionally, the loading for Coding was .28 for the AALD sample and .43 for the Kush and Watkins sample.

The standard deviations (SDs) generated from the subtest scores for both samples were then examined to determine if differences in variability of subtest

scores might account in part for these discrepancies. The SDs for Information (2.58 for the AALD sample and 2.93 for the other sample), Block Design (3.08 for the AALD sample and 3.60 for the other sample), Comprehension (2.96 for the AALD sample and 3.59 for the other sample) and Coding (3.19 for the present sample and 3.42 for the other sample) were examined. Consistently, the Kush and Watkins sample demonstrated greater variability of scores in each of these subtests. The greater variability of scores around subtest means of the Watkins and Kush sample may well account for much of the discrepancy in the four factor loading comparisons. The overall congruence of factor loadings across these two samples of African American children with disabilities provided further evidence of the reliability of the finding from the current study across other samples of African American children with disabilities.

When results of the very small Slate and Jones (1995) mixed disability sample of African American children were reviewed, support for the Verbal and Performance IQ's was indicated in the results of the principal component analyses with varimax rotations. The two subtest exceptions in rotation were Arithmetic and Coding.

In summary, the results of research involving other samples exclusively of African American children with disabilities were reexamined in light of the findings from the present study. The statistics obtained from factor analyses of all the samples of African American children consistently provided evidence to support the generalizability of the two-factor solution. Consistent with conclusions drawn from the present study, the other studies reviewed did not find statistical support for the

generalizability of the three and four factor solutions to African American children with disabilities.

Relating Findings of the Present Study to Other Ethnically Diverse Samples with Disabilities

The Konold, Kush, and Canivez study (1997) combined three samples of ethnically diverse children and subjected them to confirmatory factor analyses across five models. The first sample reflected mixed disability status. The second and third were composed of children with learning disabilities. The factor loadings were reported for the four-factor model. They were compared to the CD and AALD samples from the present study. It must be remembered in reviewing the three samples from the Konold et al. (1997) investigation that they all were ethnically dissimilar. Additionally, the composition of sample 1 with respect to disability status was very different from sample 2 and sample 3. The effects of ethnicity and disability status were not disentangled in the study.

The factor loadings for the five samples were reported in Table 33 (see Chapter IV). Examination of the factor loadings across the five samples for the subtests composing the Verbal Comprehension factor will be discussed first. The differences, which are trivial, appear to reflect merely sampling fluctuations for all loadings across all samples except for the loadings on Information on the AALD sample. This loading is much lower than that reported for the other samples. It reflects sample specific variation for that subtest.

The greatest variability in factor loadings was observed for subtests in the Freedom from Distractibility factor (FD). This factor certainly has generated the

most controversy in the literature with regard to children with exceptionality and ethnic minority samples. Perhaps the most important criticism is the fact that only two subtests define this factor. This factor is not regarded as containing sufficient measures to define itself (Gorsuch, 1983; Thorndike, 1990). For the AALD sample Arithmetic defined the factor in totality with a loading of .95, raising the question of overfactoring in that sample. Also the variance for the CD sample on the two subtests was much larger than on the other four samples. Sampling fluctuations may have accounted for some of the variance.

The variation in the Perceptual Organization factor subtests across the four samples appeared to reflect sampling fluctuation, as do the majority of the loadings for Coding for the Processing Speed factor (PS). The very high loadings of Symbol Search for the CD sample (.90) and two of the samples for the Konold et al. (1997) study (.99) suggest, however that Symbol Search is capturing in totality the latent factor. This raises the question of overfactoring in the samples. Additionally, as with FD, PS was represented by only two subtests, raising the same objections as to the adequacy of how it is defined.

The factor loadings from the maximum likelihood/direct oblique exploratory factor analysis of Kush's (1996) sample of 327 children with learning disabilities (but diverse ethnicity membership) were examined. Two large factors corresponding to Kaufman's Verbal Comprehension and Perceptual Organization factors were obtained, and a much smaller factor, corresponding to PS in the four-factor model identified by test developers of the WISC-III was identified. They were compared to the factor loadings obtained for the CD and AALD samples in the three-factor model

from a maximum likelihood/oblique partially restricted confirmatory factor analysis. Arithmetic, Digit Span, and Coding were excluded because they formed the third factor in the model used for the CD and AALD samples in contrast to the third factor in the Kush study which consisted of Coding and Symbol Search. The factor loadings for the three samples were reported in Table 34 (see Chapter IV).

The Verbal Comprehension subtests of Information, Similarities, Vocabulary, and Comprehension were examined initially. In every case the loadings from the Kush sample fall between the AALD and CD samples. They are about equidistant from both samples for Similarities, Vocabulary, and Comprehension but the factor loading of the Kush sample (.71) for the Information subtest was much closer to the CD sample (.81) than the AALD sample (.54). In the Perceptual Organization factor the factor loading of the Kush sample for Block Design (.81) was closer to the CD sample (.78) than the AALD sample (.55). By contrast the factor loading for Picture Completion for the Kush sample (.49) was smaller than it was for either the CD sample (.81) or the AALD sample (.64). The factor loading in the Kush sample for Picture Arrangement (.54) was closer to the factor loading for the AALD sample (.42) than the CD sample (.76). Additionally, the factor loading for the Kush sample for Object Assembly (.69) was closer to the AALD sample (.65) than the CD sample (.78).

The factor loadings from the Kush sample appear more similar to factor loadings from the AALD sample for some selected subtests, and more similar to factor loadings from the CD sample for other selected subtests. Several factors make it difficult to explain differences in factor loadings. They include the methodology

employed in analysis and the difference between the composition of the three-factor model for the Kush sample and the three-factor model used in the present study. Additionally, the effects of ethnicity were not disentangled in the Kush sample. The data from the Kush sample of children with learning disabilities supported the interpretation of the WISC-III using the Verbal and Performance IQ's, reflecting the two-factor model. This finding is congruent with the results of the present study. The variability in loadings underscores the need to specify sample carefully with respect to both ethnicity and disability status and wherever possible, isolate those variables in cross-sample comparisons.

Support for the Construct Validity of the WISC-III from the Current Study

Benson (1998) has proposed a framework for evaluating construct validity in a three-stage investigation. In Phase 1, psychological constructs are defined. In Phase 2, "the objective of ... [the] study is to determine the extent to which the observed variables covary among themselves, and how they covary with the intended structure of the theoretical domain" (Benson, p.13). Factor analytic techniques are used most frequently to establish the structure in Phase 2 (Benson). Messick (1995) defined the necessary condition for "structural fidelity" (p. 746) as involving consistency between "interrelations among the scored aspects of tasks and subtask performance" (p.746) and "what is known about the internal structure of the domain" (p. 746).

Phase 3 involves investigation of the meaning of the test scores. This can be established using correlations with the external criteria (Benson, 1998). Additionally, it can be established with group differentiation (Benson). "Meaning of the scores is substantiated externally by appraising the degree to which empirical relationships

with other measures---or the lack thereof--- are consistent with that meaning” (Messick, p.746). “Each stage either leads to the next in building evidence for the construct validity interpretation of test scores or suggests the previous stage should be re-evaluated” (Benson, 1998, p. 15). “Although the results of research in one stage cannot establish validity, the absence of support can undermine it” (Keith & Kranzler, 1999, p.306). Keith and Kranzler used Benson’s theoretical framework to evaluate another newer test of intelligence, concluding that it lacked structural fidelity. Keith and Kranzler interpreted this to mean “ the scaled scores ... do not reflect underlying theory” (p. 319).

The imposition of Benson’s theoretical framework upon the present study provided a meaningful way of contextualizing findings and relating them to other research that attempts to contribute evidence of the construct validity of the WISC-III. The present study, using Benson’s (1998) framework, was a study of structural fidelity, bearing in mind a difficulty in conducting WISC-III research involves the lack of a strong theoretical foundation. For that reason, the three factor models most commonly cited in the literature were investigated in this research.

The results of this study do provide support of the construct validity of the WISC-III using the two-factor structure that consists of Verbal Comprehension (composed of Information, Similarities, Vocabulary, Arithmetic, and Comprehension) and Perceptual Organization (consisting of Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding) factors. The two-factor solution provided an adequate fit across the AALD and CD samples without respecification and an excellent fit across both samples with respecification (i.e., allowing factor

covariances to be freely estimated). When Keith and Kranzler's (1999) interpretation of how structural fidelity is manifested through test scores is applied to this study, one may conclude the loadings of the subtest scores in the two-factor solution affirm the original organization of the Wechsler Scales and the use of the VIQ and PIQ in test interpretation. The consistency of the affirmation of the two-factor structure with children with and without disabilities across all versions of this test in use with children in previous research and the current study provides evidence of the structural fidelity of the two-factor model. It remains for further research on how well the subtest scores of the WISC-III predict achievement in samples of African American children with learning disabilities, a Phase 3 future study, to establish how well they provide claims of evidence of external validity.

The three and four-factor solutions for the AALD sample in the current study did not substantiate the structural fidelity of these models for this sample. The absence of support is reflected in the results of the partially restricted and fully restricted factor analyses involving the AALD sample. Weaker claims of evidence both with regard to the Freedom from Distractibility factor in the three-factor model and the use of Index scores in the four-factor model are evident in model evaluation statistics in the three and four-factor solutions.

The Use of Index Scores with African American Children with Learning Disabilities

Results of the investigation of the generalizability of the factor structure of the WISC-III across the two-factor, three-factor, and four-factor models did not support the use of Index scores with this sample of African American school-age children with learning disabilities. The model evaluation statistics generated for the four-factor

model reflected a marginal fit when the factor structure of samples without disabilities were imposed on the AALD sample and an unacceptable fit when the factor structure of the CD sample was imposed on the AALD sample.

In addition to considering the results of the factor analyses, which strongly support the two-factor model alone with this sample, examination of the mean subtest scores revealed that the three subtests reflecting lowest scores are Information, Vocabulary, and Arithmetic. If index scores were used for the AALD sample, the low scores on Information and Vocabulary would suppress the Index score for the Verbal Comprehension index. The low score on Arithmetic would suppress the index score for the Freedom from Distractibility index. Because of these effects, the index scores would not provide a more reliable estimate of ability than the traditional global estimates of intelligence (i.e., FSIQ, VIQ, and PIQ).

An additional finding emerged related to PS in the four-factor model. In the AALD sample, Coding emerged as the subtest with the highest mean score. Although the finding may be sample specific, it is buttressed by the mean PS Index computed for the portion of the AALD sample for which Symbol Search scores were reported. The PS Index score of 89.23 was the highest index score generated for the sample. This finding is congruent with Prifitera, Weiss, and Saklofske's (1988) report that it is the highest mean index score for the WISC-III African American and Hispanic samples as well, although the lowest mean index score for the Caucasian sample.

Furthermore, the loading for Symbol Search in the AALD sample was higher (.71) than the loading for Coding (.51) in the four-factor model. This finding for the

AALD sample differs from Carroll's (1993b) findings in his factor analysis of the WISC-III standardization sample. Carroll reported Processing Speed was characterized largely by Coding and minimally by Symbol Search in the WISC-III standardization sample. For the AALD sample Symbol Search appears to characterize the factor to a greater degree than Coding.

Examination of the loading of Symbol Search in the four-factor model for the CD sample revealed an even higher loading (.90). The loading for Symbol Search in the four-factor model for the CN sample was almost identical (.91). The loading for Symbol Search in the four-factor model for the AAN sample was .80. These findings, in contrast to Carroll's, suggest Symbol Search defines Processing Speed to a greater degree in all samples in this study.

How useful would this index be in general assessment practice? Data from the WISC-III manual (Wechsler, 1991) indicate PS is a poor correlate of academic achievement. Use of the WISC-III by school psychologists is heavily tied to their work in determining eligibility of students for special services. In that process they try to identify strengths and weaknesses relevant to academic progress. PS appears to offer little additional information that would be meaningful in the assessment process in relation to children designated with learning disabilities who do not evidence other kinds of neurological impairments.

The Use of Index Scores in This Sample of African American Children with Disabilities

Glutting, Konold, McDermott, Kush, and Watkins (1996), in their analyses of a large mixed sample across six states reported about one third of the test protocols

included Digit Span and Symbol Search, both optional subtests required to generate Index scores. Other researchers (Blumberg, 1995; Ward, Ward, Hatt, Young, & Mollner, 1995) confirmed this finding. In the present study the score for Symbol Search was reported in less than 20 % of the protocols. Digit Span, by contrast, was administered in about 59 % of the protocols. Consistent with previous reports, it appears Symbol Search is not being used with great frequency in the general assessment practice of school psychologists.

Substituting Symbol Search for Coding in Calculating Global Scores

The findings of the present study preliminarily suggest it might be appropriate to substitute Symbol Search for Coding in the calculation of the Performance and Full Scale Intelligence Quotients in testing African American children with learning disabilities. Neither Coding nor Symbol Search correlated highly with the other subtests in the Performance factor for the two-factor model but Symbol Search has a much higher factor loading on PS than Coding in the four-factor model. This suggests it is a better predictor of that factor than Coding and therefore possibly a better subtest to use in calculating the Performance and Full Scale IQs.

Limitations of This Study

The African American sample with learning disabilities in this study was drawn from only two of the four geographic regions specified in the 1988 United States Bureau of Census report (i.e., the Northeast and the Southeast). Additionally, this sample included only urban and suburban children. Application to children living in other geographical regions and children living in rural areas remains unclear.

Data gathered from the sample of school-age children did not reflect the socioeconomic status of African Americans in proportions equivalent to those noted in the 1988 United States Census. This was not a stratified sample with respect to socioeconomic status. In fact, a higher proportion of the children in the school districts from which the sample was drawn were eligible for free and reduced price lunches than is the case in the general population.

The effect of including greater numbers of children living in poverty offered one possible advantage. In actual practice, larger numbers of African American children who have low socioeconomic status are designated as having learning disabilities. The AALD sample probably was more similar to the population of African American children being tested in schools than the AAN sample from the WISC-III standardization sample. A stratified sample, however, would provide stronger evidence of the generalizability of the factor structure across all socioeconomic levels.

Archival data were collected from only two states in constructing the AALD sample. It was not possible to determine if there were variations within and among the districts in the way their personnel interpreted and applied state regulations and federal regulations to determine eligibility for special education services. Additionally, it is not known the degree to which generalizability of findings from the current sample to other samples of African American children with learning disabilities would be affected by differences in state regulations in other states and the interpretation thereof by members of the Committees on Special Education in those states. Differences in criteria for designation conceivably could have an effect on the

composition of the sample and be reflected in the factor structure generated from the covariance matrix of the test scores.

Recommendations for Future Research

The present study, in summary, extends the work of Allen and Thorndike (1995a, 1995b) using their cross validation of covariance structure matrices methodology (CVCSM) to investigate the generalizability of the factor structure of the WISC-III to African American children with learning disabilities. The use of CVCSM enabled this study to disaggregate the effects of ethnicity and disability status across the CN, AAN, and CD samples of the WISC-III standardization sample and the AALD sample collected for this study. It provides claims of evidence that the WISC-III is invariant across ethnicity and disability status in the two-factor model that reflects Wechsler's original conceptualization of the Wechsler Scales.

The relatively small sample size in this study utilized in the cross-age restricted factor analysis of the 12 subtests which are purported to underlie the four-factor structure of the WISC-III standardization sample raises questions about the generalizability of results in the four-factor model, however. Further studies will need to be conducted using CVCSM with other large samples of African American children with learning disabilities to determine if results can be replicated. Additionally, the sample size for the AALD sample in the four-factor model was insufficient to allow for cross validation. Cross validation would provide stronger claims of evidence of factor structure reliability across samples and needs to be undertaken.

The AALD sample was a heterogeneous sample of children with learning disabilities. It was not possible to obtain information about the individual types of learning disabilities that define the disability status of each child included in the sample. Therefore, the degree of inter-subtest variability present in the sample was unknown. It would be interesting to study the factor structure in samples of African American children with more homogeneously defined specific learning disabilities (i.e., dyslexia, dysgraphia, acalculia, etc.) to evaluate whether factor structure reliability would be replicated in the same way it was in the present study.

REFERENCES

Alfonso, V. F., Oakland, T. D., LaRocca, R., & Spanakos, A. (in press). The course on individual cognitive assessment. School Psychology Review.

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

Allen, S. R., & Thorndike, R. M. (1995b). Stability of the WAIS-R and WISC-III factor structure using cross-validation of covariance structures. Journal of Clinical Psychology, 51, 3-20.

Anastasi, A. (1976). Psychological testing (4th ed.). New York: Macmillan.

Anastasi, A. (1982). Psychological testing (5th ed.). New York: Macmillan.

Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan.

Anastasi, A., & Urbina, S. (1997). Psychological testing. (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Artiles, A. J., Trent, S. C., & Kuan, L. (1997). Learning disabilities empirical research on ethnic minority students: An analysis of 22 years of studies published in selected refereed journals. Learning Disabilities Research and Practice, 12, 82-91.

Bell, W. M. (1994). A confirmatory factor analysis of the Wechsler Intelligence Scale for Children - Third Edition on children with learning disabilities. Unpublished doctoral dissertation, Mississippi State, Starkville.

Bennett, R. E. (1983). Research and evaluation priorities for special education assessments. Exceptional Children, 50, 110-115.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. Educational Measurement Issues and Practices, 17, 10-17.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 31, 419-456.

Bentler, P. M. (1988). Causal modeling via structural equation systems. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (2nd ed., pp. 317-335). New York: Plenum.

Bentler, P. M. (1989). EQS structural equations program manual. Los Angeles: BMDP Statistical Software.

Bentler, P. M. (1990a). Fit indexes, Lagrange Multipliers, constraint changes and incomplete data in structural models. Multivariate Behavioral Research, 25, 163-172.

Bentler, P. M. (1990b). Comparative Fit Indexes in Structural Models. Psychological Bulletin, 107, 238-246.

Bentler, P. M. (1995). EQS structural equations program manual. Encino, CA: Multivariate Software, Inc.

Bentler, P. M., & Wu, E. J., (1995). EQS for Windows: User's guide. Encino, CA: Multivariate Software, Inc.

Blaha, J., & Wallbrown, F. H. (1984). Hierarchical analyses of the WISC and WISC-R: Synthesis and clinical implications. Journal of Clinical Psychology, 40, 556-571.

Blaha, J., & Wallbrown, F. H. (1996). Hierarchical factor structure of the Wechsler Intelligence Scale for Children-III. Psychological Assessment, 8, 214-218.

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

Borgas, K. (1999, Winter). Intelligence theories and psychological assessment: Which theory of intelligence guides your interpretation of intelligence test profiles? The School Psychologist, *53*, 24-26.

Braden, J. P. (1987). A comparison of regression and standard score discrepancy methods for learning disabilities identification: Effects of racial representation. Journal of School Psychology, *25*, 23-29.

Braden, J. P. (1997). The practical impact of intellectual assessment issues. School Psychology Review, *26*, 242-248.

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? Psychological Bulletin, *107*, 260-273.

Broman, S. H., Nichols, P. L., & Kennedy, W. A. (1975). Preschool IQ: Prenatal and early developmental correlates. Hillsdale, NJ: Erlbaum.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. Developmental Psychology, *22*, 723-742.

Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since Bias in Mental Testing. School Psychology Quarterly, *14*, 208-238.

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics (pp. 99-136). Washington, DC: American Psychological Association.

Byrne, B. M. (1989). Multigroup comparisons and the assumptions of construct validity across groups: Methodological and substantive issues. Multivariate Behavioral Research, *24*, 503-523.

Byrne, B. M. (1994a). Testing for the factorial validity, replication, and invariance of a measurement instrument: A paradigmatic application based on the Maslach Burnout Inventory. Multivariate Behavioral Research, 29, 289-311.

Byrne, B. M. (1994b). Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications and programming. Thousand Oaks, CA: Sage.

Canivez, G. L., & Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children-Third Edition. Psychological Assessment, 10, 285-291.

Carroll, J. B. (1993a). Human cognitive abilities: A survey of factor-analytic studies. New York: Cambridge University Press.

Carroll, J. B. (1993b). What abilities are measured by the WISC-III? [Monograph Series: Advances in Psychoeducational Assessment: Wechsler Intelligence scale for Children: Third Edition]. Journal of Psychoeducational Assessment, 134-143.

Chalfant, J. C. (1989). Learning disabilities: Policy issues and promising approaches. American Psychologist, 44, 392-398.

Chattin, S. H., & Bracken, B. A. (1989). School psychologists' evaluations of the K-ABC, McCarthy Scales, Stanford-Binet IV, and WISC-R. Journal of Psychoeducational Assessment, 7, 112-130.

Clarizio, H. F. (1978). Nonbiased assessment of minority group children. Measurement and Evaluation in Guidance, 11, 106-113.

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

Cole, N. S., & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), Educational measurement, (3rd ed., pp. 201-219). New York: Macmillan.

Cone, T. E., & Wilson, L. R. (1981). Quantifying a severe discrepancy: A critical analysis. Learning Disability Quarterly, *4*, 359-371.

Cox, T. (1983). Cumulative deficit in culturally disadvantaged children. British Journal of Educational Psychology, *53*, 317-326.

Cronbach, L. J. (1970). Essentials of psychological testing (3rd ed.). New York: Harper & Row.

Crown, P. J. (1970). The effects of race of examiner and standard vs. dialect administration of the Wechsler Preschool and Primary Scale of Intelligence on the performance Negro and White Children. Dissertation Abstracts International, *32*, 232A-233A. (University Microfilms No. 71-18, 356).

DiGangi, S. A., & Faykus, S. P. (1993). Assessment for effective instruction. In H. Boone Vance (Ed.), Best practices in assessment for school and clinical settings (pp. 271-305). Brandon, VT: Clinical Psychology Publishing Co., Inc.

Donders, S. J. (1996). Cluster subtypes in the WISC-III standardization sample: Analysis of factor index scores. Psychological Assessment, *8*, 312-318.

Duncan, G. (1994). Families and neighbors as sources of disadvantage in the schooling decisions of white and black adolescents. American Journal of Education, *105*, 20-53.

- Duncan, G., & Rodgers, W. (1988). Longitudinal aspects of childhood poverty. Journal of Marriage and the Family, *50*, 1007-1021.
- Durrant, J. E. (1994). A decade of research on learning disabilities: A report card on the state of the literature. Journal of Learning Disabilities, *27*, 25-33.
- Eggenbeen, D., & Lichter, D. (1991). Race, family structure, and changing poverty among American children. American Sociological Review, *56*, 801-817.
- Elliott, C. D. (1990). Introductory and technical handbook for the Differential Ability Scales. San Antonio: Psychological Corporation.
- Elliott, S. N., Piersel, W. C., Argulewicz, E. N., Gutkin, T. B., & Galvin, G. A. (1985). Three year stability of WISC-R IQ's for handicapped children from three racial/ethnic groups. Journal of Psychoeducational Assessment, *3*, 233-244.
- Esters, I. G., Ittenbach, R. F., & Han, K. (1997). Today's IQ tests: Are they really better than their historical predecessors? School Psychology Review, *26*, 211-224.
- Eysenck, H. J. (1991). Race and intelligence: An alternative hypothesis. Mankind Quarterly, *32*, 123-125.
- Fan, X., Willson, V. L., & Kapes, J. T. (1994, August). Ethnic group's representation in test construction sample and test bias. Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.
- Fan, X., Willson, V. L., & Kapes, J. T. (1996). Ethnic group representation in test construction samples and test bias: The standardization fallacy revisited. Educational and Psychological Measurement, *56*, 365-381.

Fan, X., Willson, V. L., & Reynolds, C. (1995). Assessing the similarity of the factor structure of the K-ABC for African American and White children. Journal of Psychoeducational Assessment, *13*, 120-131.

Federal Register. (1999, March). Final regulations for the assistance to states for the education of children with disabilities and the early intervention program for infants and toddlers with disabilities. Department of Education, pp.12406-12672.

Figueroa, R. A. (1990). Assessment of linguistic minority group children. In C. R. Reynolds & R. W. Kamphaus, (Eds.), Handbook of Psychological and Educational Assessment of Children: Vol. 1 Intelligence and Achievement. New York: Guilford.

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

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. Psychological Bulletin, *95*, 29-51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin, *101*, 171-191.

Frame, R. (1979, September). Diagnoses related to school achievement, client's race and socio-economic status. Paper Presented at the Annual Meeting of the American Psychological Association, New York.

Frankenberger, W., & Harper, J. (1987). States' criteria and procedures for identifying learning disabled children: A comparison of 1981-82 and 1985/86 guidelines. Journal of Learning Disabilities, *20*, 118-121.

Gaddes, W. H., & Edgell, D. (1993). Learning disabilities and brain function (3rd ed.). New York: Springer-Verlag.

Gass, C. S., & Demsky, Y. I. (1998). Factor analysis of the WISC-R (Spanish Version) at 11 age levels between 6 1/2 and 16 1/2 years. Journal of Clinical Psychology, 54, 109-113.

Glutting, J., McDermott, P. A., & Konold, T. R. (1997). Ontology, structure, and diagnostic benefits of a normative subtest taxonomy from the WISC-III standardization sample. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 349-372). New York: Guilford Press.

Glutting, J., McDermott, P. A., Prifitera, A., & McGrath, E. (1994). Core profile types for the WISC-III and WIAT: Their development and application in identifying multivariate IQ - achievement discrepancies. School Psychology Review, 23, 619-639.

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

Granier, M. J., & O'Donnell, L. (1991, August). Children's WISC-III scores: Impact of parent education and home environment. Poster session presented at the annual meeting of the American Psychological Association, San Francisco, CA.

Graziano, W. G., Varga, P. E., & Levy, J. C. (1982). Race of examiner effects and the validity of intelligence tests. Review of Educational Research, 52, 469-497.

Gregg, N. (1995, December). Assessment and intervention for African Americans with learning disabilities. Communiqué, 17.

Griffith, J. (1994, August). Presentation of racial groups in special education:

Some conceptual and methodological considerations. Poster session presented at the annual meeting of the American Psychological Association, Los Angeles, CA.

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, 227-231.

Hale, R. L. (1983). An examination for construct bias in the WISC-R across socioeconomic status. Journal of School Psychology, 21, 153-156.

Harmon, H. H. (1967). Modern factor analysis (2nd ed.). Chicago: University of Chicago Press.

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

Harry, B. (1994). The disproportionate representation of minority students in special education: Theories and recommendations. Alexandria, VA: National Association of State Directors of Special Education.

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

Hill, T. D., Reddon, J. R., & Jackson, D. N. (1985). The factor structure of the Wechsler Scales: A brief review. Clinical Psychology Review, 5, 287-306.

Hishinuma, E. S., & Yamakawa, R. (1993). Construct and criterion-related validity of the WISC-III for exceptional students and those who are "at risk". [Monograph Series: Advances in Psychoeducational Assessment: Wechsler Intelligence Scale for Children: Third Edition]. Journal of Psychoeducational Assessment, 94-104.

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

Hutton, J. B., Dubes, R., & Muir, S. (1992). Assessment practices of school psychologists: Ten years later. School Psychology Review, 21, 271-284.

Individuals with Disabilities Education Act, 34 C.F.R. pt. 300 (1991).

Individuals with Disabilities Education Act, 20 U.S.C. ch.33, SS 1400-1485 (1999).

Ittenbach, R. F., Esters, I. G., & Wainer, H. (1997). The history of test development. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 17-31). New York: Guilford Press.

Jargowsky, P. (1994). Ghetto poverty among Blacks in the 1980s. Journal of Policy Analysis and Management, 13, 288-310.

Jencks, C. (1972). Inequality: A reassessment of the effect of family and schooling in America. New York: Harper & Row.

Jensen, A. R. (1977). Cumulative deficit in IQ of blacks in the rural south. Developmental Psychology, 13, 184-191.

Jensen, A. R. (1998). The g factor: The science of mental ability. Westport, CT: Praeger.

Jensen, A. R., & Reynolds, C. R. (1982). Race, social class and ability patterns on the WISC-R. Personality and Individual Differences, 3, 423-438.

Jones, D. R., & James, S. (1993). Best uses of the WISC-III. In H. Booney Vance (Ed.), Best practices in assessment for school and clinical setting (pp. 231-269). Brandon, VT: Clinical Psychology Publishing Co., Inc.

Joreskog, K. G., & Sorbom, D. (1981). LISREL V: Analysis of linear structural relations by the method of maximum likelihood. Chicago: International Educational Services.

Joreskog, K. G., & Sorbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Hillsdale, NJ: Lawrence Erlbaum Associates.

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. (1993). Clinical assessment of children's intelligence: A handbook for professional practice. Needham Heights, MA: Allyn and Bacon.

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

Kamphaus, R. W., Petoskey, M. D., & Morgan, A.W. (1997). A history of intelligence test interpretation. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 32-47). New York: Guilford Press.

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. Multivariate Behavioral Research, 25, 137 - 155.

Karnes, F. A., & Brown, K. E. (1980). Factor analysis of the WISC-R for the gifted. Journal of Educational Psychology, *72*, 197-199.

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*, 135-147.

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

Kaufman, A. S. (1982). The impact of WISC-R research for school psychologists. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology (pp. 156-177). New York: Wiley.

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

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

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

Kaufman, A. S., & Kaufman, N. (1985). Kaufman Test of Educational Achievement. Circle Pines, MN: American Guidance Services Inc.

Kavale, K. A., & Forness, S. R. (1984). A meta-analysis of the validity of Wechsler scale profiles and recategorizations: Patterns or parodies? Learning Disabilities Quarterly, *7*, 136-156.

Keith, T. Z., & Reynolds, C. R., (1990). Measurement and design issues in child assessment research. In C. R. Reynolds & R. W. Kamphaus (Eds.), Handbook

of psychological & educational assessment of children: Intelligence & achievement (pp. 29-61). New York: Guilford.

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

Keith, T. Z., & Kranzler, J. H. (1999). The absence of structural fidelity precludes construct validity: Rejoinder to Naglieri on what the Cognitive Assessment System does and does not measure. School Psychology Review, 28, 303-321.

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

Klerman, L. V., & Parker, M. B. (1991). Alive and well? A research and policy review of health programs for poor young children New York: National Center for Children in Poverty (Columbia University School of Public Health).

Konold, T. R., Glutting, J. J., McDermott, P. A., Kush, J. C., & Watkins, M. M. (1999). Structure and diagnostic benefits of a normative subtest taxonomy developed from the WISC-III standardization sample. Journal of School Psychology, 37, 29-48.

Konold, T. R., Kush, J. C., & Canivez, G. L. (1997). Factor replication of the WISC-III in three independent samples of children receiving special education. Journal of Psychoeducational Assessment, 15, 123-137.

Kranzler, J. H. (1997). Educational and policy issues related to the use and interpretation of intelligence tests in the schools. School Psychology Review, 26, 150-162.

Kush J. C. (1996). Factor structure of the WISC-III for students with learning disabilities. Journal of Psychoeducational Assessment, 14, 32-40.

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

Kush, J. C., Watkins, M. W., Ward, T. J., Ward, S. B., Canivez, G. L., & Worrell, F.C. (in press). Construct validity of the WISC-III for Black students referred for psychological evaluation.

Larry P. et al. v. Riles (1979, October). United States District Court for the Northern District of California, C-71-2270 RFP.

Larry P. v. Riles (1986). United States District Court for the Northern District of California, C-71-2270 RFP. [Order Modifying Judgment.]

Leamer, E. E. (1978). Specification searches: Ad hoc inference with non-experimental data. New York: John Wiley & Sons.

Lindsey, J. (1967). The factorial organization of intelligence in children as related to the variables of age, sex, and subculture. Dissertation Abstracts, 27, 3664A-3665A. (University Microfilms No. 67-3567)

Loderquist-Hansen, S., & Barona, A. (1994, August). Factor structure of the Wechsler Intelligence Scale for Children-III for Hispanic and non-Hispanic children

with learning disabilities. Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.

Loehlin, J. C. (1992). Latent variable models: An introduction to factor, path, and structural analysis (3rd ed.). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Long, J. S. (1983). Covariance structure models: An introduction to LISREL. Beverly Hills, CA: Sage.

Lucas, A., Morley, R., Cole, T.J., Lister, G., & Leeson-Payne, C. (1992). Breast milk and subsequent intelligence quotient in children born preterm. Lancet, 339, 261-264.

Lutey, C., & Copeland, E. P. (1982). Cognitive assessments of the school-age child. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology (pp. 121-155). New York: John Wiley & Sons.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. Psychological Bulletin, 100, 107-120.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. Psychological Bulletin, 111, 490-504.

MacCallum, R. C., Roznowski, C. M., & Reith, J. V. (1994). Alternative strategies for cross-validation strategies of covariance structure models. Multivariate Behavioral Research, 29, 1-32.

MacMillan, D. L., Gresham, F. M., & Bocian, K. M. (1998a). Discrepancy between definitions of learning disabilities and school practices: An empirical investigation. Journal of Learning Disabilities, 31, 314-326.

MacMillan, D. L., Gresham, F. M., & Bocian, K. M. (1998b). Curing mental retardation and causing learning disabilities: Consequences of using various WISC-III IQs to estimate aptitude of Hispanic students. Journal of Psychoeducational Assessment, 16, 36-54.

MacMillan, D. L., Gresham, F. M., Lopez, M. F., & Bocian, K. M. (1996). Comparisons of students nominated for prereferral interventions by ethnicity and gender. The Journal of Special Education, 30, 133-151.

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

Macmann, G. M., & Barnett, D. W. (1994). Structured analysis of correlational factors: Lessons from the verbal-performance dichotomy of the Wechsler scales. School Psychology Quarterly, 9, 161-197.

Mardell-Czudnowski, C., & Burke, L. J. (1991, August). Testing: Who, what, and why and whom a decade later. Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

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, 391-410.

Matarazzo, J. S., & Herman, D. O. (1985). Clinical uses of the WAIS-R: Base rates of differences between VIQ and PIQ in the WAIS-R standardization sample. In B.B. Wolman (Ed.), Handbook of Intelligence: Theories, Measurements and Applications (pp. 899-933). New York: John Wiley & Sons.

Matuszek, P., & Oakland, T. (1979). Factors influencing teachers' and psychologists' recommendations regarding special class placement. Journal of School Psychology, 17, 116-125.

McBrayer, K. F., & Garcia, S. (1996, April). Asian/Pacific American students in special education, A comparative analysis of service incidence trends. Paper presented at the annual meeting of the Council for Exceptional Children, Orlando, FL.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. Psychological Bulletin, 107, 247-255.

McLeskey, J., Waldron, N. L., & Wornhoff, S. A. (1990). Factors influencing the identification of Black and White students with learning disabilities. Journal of Learning Disabilities, 23, 362-366.

McLeskey, J., & Waldron, N. L. (1991). Identifying students with learning disabilities: The effect of implementing statewide guidelines. Journal of Learning Disabilities, 24, 501-506.

McLoyd, V. V. (1990). The impact of economic hardship on black families and children: Psychological distress, parenting, and socioeconomic stress, parenting and socio-emotional development. Child Development, 61, 311-346.

McLoyd, V. (1998). Socioeconomic disadvantage and child development. American Psychologist, 53, 185-204.

Mercer, C. D., Jordan, L., Allsopp, D. H., & Mercer, A. R. (1996). Learning disabilities definitions and criteria used by state education departments. Learning Disability Quarterly, 19, 217-232.

Mercer, J. (1979). System of multicultural pluralistic assessment technical manual. New York: The Psychological Corporation.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. American Psychologist, *30*, 955-966.

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, *35*, 1012-1027.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. Psychological Bulletin, *89*, 575-588.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-90). New York: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific enquiry into score meaning. American Psychologist, *50*, 741-749.

Meyers, C. E., MacMillan, D., & Yoshida, R. (1978). Validity of psychologists identification of EMR students in the perspective of the California decertification experience. Journal of School Psychology, *16*, 3-15.

Misra, G. (1983). Deprivation and development: A review of Indian studies. Indian Educational Review, 12-32.

Moore, C. L., & Retish, P. M. (1974). Effect of the examiner's race on Black children's Wechsler Preschool and Primary Scale of Intelligence IQ. Developmental Psychology, *10*, 672-676.

Morison, P., White, S. H., & Feuer, M. J. (Eds.). (1996). The use of IQ tests in special education decision making and planning: Summary of two workshops.

Board on Testing and Assessment. Commission on Behavioral and Social Sciences and Education. National Research Council. Washington, DC: National Academy Press.

Most, R. B., & Zeidner, M. (1995). Constructing personality and intelligence instruments: Methods and issues. In C. R. Reynolds & R. T. Brown (Series Eds.) & D. H. Salofse & M. Zeidner (Vol. Eds.), International handbook of personality and intelligence: Perspectives on individual differences (pp. 475-503). New York: Plenum Press.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, *105*, 430-445.

Naglieri, J. A. (1981). Factor structure of the WISC-R for children identified as learning disabled. Psychological Reports, *49*, 891-895.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A.W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. American Psychologist, *51*, 77-101.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Oakland, T., & Parmelee, R. (1985). Mental measurement of minority group children. In B. B. Wolman (Ed.), Handbook of intelligence: Theories, measurements and applications (pp. 699-736). New York: John Wiley & Sons.

Oakman, S., & Wilson, B. (1988). Stability of WISC-R intelligence scores: Implications for 3-year evaluations of learning disabled students. Psychology in the Schools, 25, 118-120.

O'Donnell, L., Granier, M. J., & Dersh, J. J. (1991, August). Does handedness affect children's coding performance on the WISC-III? Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

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

Oswald, D. P., Coutinho, M. J., Best, A. M., & Singh, N. N. (1999). Ethnic representation in special education: The influence of school-related economic and demographic variables. The Journal of Special Education, 32, 194-206.

Ownby, R. L., & Carmin, C. C. (1994, August). Confirmatory factor analysis of the WISC-III: Age-related changes in factor structure and alternative factor models. Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.

Ownby, R. L. & Matthews, C. G. (1985). On the meaning of the WISC-R third factor: Relations to selected neuropsychological measures. Journal of Consulting and Clinical Psychology, 53, 531-534.

Petersen, C. R. & Hart, D. H. (1979). Factor structure of the WISC-R for a Clinic-Referred Population and specific subgroups. Journal of Counseling and Clinical Psychology, 47, 643-645.

Prifitera, A., & Dersh, J. (1993). Base rates of WISC-III diagnostic subtest patterns among normal, learning-disabled, and ADHD samples. Journal of

Psychoeducational Assessment monograph series, Advances in psychological assessment: Wechsler Intelligence Scale for Children--Third Edition, 43-55.

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

Pryzwansky, W., Nicholson, C. L., & Uhl, N. P. (1974). The influence of examiner race on the cognitive functioning of urban and rural children of different races. Journal of School Psychology, 12, 2-7.

Quay, L. C. (1974). Language dialect, age, and intelligence-test performance in disadvantaged Black children. Child Development, 45, 463-468.

Reschly, D. J. (1978). WISC-R Factor structures among Anglos, Blacks, Chicanos, and Native-American Papagos. Journal of Clinical and Consulting Psychology, 46, 417-422.

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

Reschly, D. J. (1997). Diagnostic and treatment utility of intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp.437-456). New York: Guilford Press.

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

Reynolds, C. R. (1995). Test bias and the assessment of intelligence and personality. In D. H. Saklofske & M. Zeidner (Eds.), International handbook of personality and intelligence (pp. 545-573). New York: Plenum Press.

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

Reynolds, C. R., & Kaiser, S. (1990). Test bias in psychological assessment. In T. B. Gutkin & C. R. Reynolds (Eds.), The handbook of school psychology (2nd ed., pp. 487-525). New York: Wiley.

Reynolds, C. R., & Kaufman, A. S. (1985). Clinical assessment of children's intelligence with the Wechsler Scales. In B.B. Wolman (Ed.), Handbook of intelligence: Theories, measurements and applications (601-661). New York: John Wiley & Sons.

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

Riccio, C. A., Cohen, M. J., Hall, J., & Ross, C. M. (1997). The third and fourth factors of the WISC-III: What they don't measure. Journal of Psychoeducational Assessment, 15, 27-39.

Robertson, P., Kushner, M. I., Starks, J., & Drescher, C. (1994). An update of participation rates of culturally and linguistically diverse students in special education: The need for a research and policy agenda. The Bilingual Special Education Perspective, 14, 3-9.

Roid, G. H., Prifitera, A., & Weiss, L. G. (1993). Replication of the WISC-III factor structure in an independent sample. [Monograph series: Advances in psychoeducational assessment: Wechsler Intelligence Scale for Children: Third Edition] Journal of Psychoeducational Assessment, 7-21.

Roid, G. H., & Worrall, W. (1997). Replication of the Wechsler Intelligence Scale for Children-Third Edition four-factor model in the Canadian normative sample. Psychological Assessment, 9, 512-515.

Rowe, D. (1994). No more than skin deep. American Psychologist, 49, 215-216.

Sabatino, D. A., Spangler, R. S., & Vance, H. B. (1995). The relationship between the Wechsler Intelligence Scale for Children-Revised and the Wechsler Intelligence Scale for Children-III scales and subtests with gifted children. Psychology in the Schools, 32, 18-23.

Saco-Pollitt, C., Pöllitt, E., & Greenfield, D. (1985). The cumulative deficit hypothesis in the light of cross-cultural evidence. International Journal of Behavioral Development, 8, 75-97.

Salkind, N. J. 1994. Child development (7th ed.). Fort Worth, TX: The Harcourt Press.

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

Sapp, G.L., Chissom, B., & Graham, E. (1985). Factor analysis of the WISC-R for gifted students: A replication and comparison. Psychological Reports, 57, 947-951.

Sattler, J. M. (1981). Intelligence tests on trial: an "interview" with judges Robert F. Peckham and John F. Grady. The Journal of School Psychology, 19, 359-369.

Sattler, J. M. (1982). Assessment of children's intelligence and special abilities (2nd ed.). Boston: Allyn and Bacon, Inc.

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

Semler, I., & Iscoe, I. (1966). Structure of intelligence in Negro and White children. Journal of Educational Psychology, 57, 326-336.

Shaw, S. R., Swerdlik, M. E., & Laurent, J. (1993). The Wechsler Intelligence Scale for Children - Third Edition [Monograph series: Advances in psychoeducational assessment: Wechsler Intelligence Scale for Children: Third Edition]. Journal of Psychoeducational Assessment, 1, 151-160.

Sherman, M., & Key, C. V. (1932). The intelligence test scores of isolated mountain children. Child Development, 3, 279-290.

Shuey, A. M. (1966). The testing of Negro intelligence. (2nd ed.). New York: Social Sciences Press.

Silver, A. A., & Hagin, R. A. (1990). Disorders of learning in childhood. New York: John Wiley & Sons.

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

Silvia, E. S. (1988). Effects of sampling error and model misspecification on goodness-of-fit indices for structural equation models. Unpublished doctoral dissertation, Ohio State University, Columbus, Ohio.

Slate, J. R., & Jones, C. H. (1995). Preliminary evidence of the validity of the WISC-III for African American students undergoing special education evaluation. Educational and Psychological Measurement, 55, 1067-1074.

Smith, D. D., Deshler, D., Halahan, D., Lovitt, T., Robinson, S., Vorees, J., & Ysseldyke, J. (1984). Minimum standards for the description of subjects in learning disabilities research reports. Learning Disability Quarterly, 7, 221-225.

Smith, M. D., & Rogers, C. M. (1978). Reliability of standardized assessment instruments when used with learning disabled children. Learning Disability Quarterly, 3, 23-31.

Speece, D. (1994). The role of classification in learning disabilities. In S. Vaughn & C. Bos (Eds.), Research issues in learning disabilities: Theory, methodology, assessment, and ethics (pp. 69-82). New York: Springer-Verlag.

Stavrou, E., & Flanagan, R. (1996, March). The stability of the WISC-III scores in learning disabled children. Paper presented at the Annual Convention of the National Association of School Psychologists, Atlanta, GA.

Stewart, K. J., & Moely, B. E. (1983). The WISC-R third factor: What does it mean? Journal of Consulting and Clinical Psychology, 46, 869-872.

Stone, B. J. (1992). Joint confirmatory factor analyses of the DAS and WISC-R. Journal of School Psychology, 30, 185-195.

Suzuki, L. A. (1992). Ethnic ability patterns on the Wechsler Intelligence Scale for Children - Revised: Relationship to theories of intelligence and cognition. Unpublished doctoral dissertation. University of Nebraska, Lincoln.

Tanaka, J. S. (1990). Towards the second generation of structural modeling. Multivariate Behavioral Research, 25, 187 - 191.

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. Educational and Psychological Measurement, 57, 5-19.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. Educational and Psychological Measurement, 56, 197-208.

Thorndike, R. L. (1963). The concepts of over-and under-achievement. New York: Teachers College Columbia University.

Thorndike, R. M. (1978). Correlational procedures for research. New York: Gardner Press.

Thorndike, R. M. (1990). Would the real factors of the Stanford-Binet Fourth Edition please come forward? Journal of Psychoeducational Assessment, 8, 412-435.

Thorndike, R. M. (1992, March). Intelligence tests: What we have and what we should have. Paper presented at the meeting of the National Association of School Psychologists, Nashville, TN.

Thorndike, R. M. (1997). The origins of intellectual assessment. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 3-16). New York: Guilford Press.

Torgesen, J. K., & Dice, C. (1980). Characteristics of research on learning disabilities. Journal of Learning Disabilities, 13, 531-535.

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

U.S. Bureau of the Census (1989). Statistical abstract of the United States: 1988 (109th ed.). Washington, DC: Government Printing Office.

U.S. Bureau of the Census. (1994). Statistical abstract of the United States:1994. Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census. (1996). Statistical abstract of the United States: 1996. Washington, DC: U.S. Government Printing Office.

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

Vance, H. R., Huelsman, C. B., Jr., & Wherry, R. J. (1976). The hierarchial factor structure of the Wechsler Intelligence Scale for Children as it relates to disadvantaged White and Black children. Journal of General Psychology, 95, 287-293.

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

Vandivier, P. L., & Vandivier, S. S. (1979). A nonbiased assessment of intelligence testing. The Educational Forum, 44, 97-108.

Vernon, P. (1950). The structure of human abilities. London: Methuen.

Vernon, P. A., Jackson, D. N., & Messick, S. (1988). Cultural influence on patterns of abilities in North America. In S. H. Irvine and J. W. Berry (Eds.), Human abilities in cultural context. New York: Cambridge University Press.

Ward, S. B., Ward, T. J., Hatt, C. V., Young, D. L., & Mollner, N. R. (1995). The incidence of the ACID, ACIDS, and SCAD profiles in a referred population. Psychology in the Schools, 32, 267-276.

Ward, T. J., Ward, S. B., Glutting, J. J., & Hatt, C. V. (1999). Exceptional LD profile types for the WISC-III and WIAT. School Psychology Review, 28, 629-643.

Watkins, M. W., & Kush, J. C. (1994). Wechsler subtest analysis: The right way, the wrong way, or no way? School Psychology Review, 23, 640-651.

Wechsler, D. (1939). The measurement of adult intelligence. Baltimore: Williams & Wilkins.

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

Wechsler, D. (1955). Manual for the Wechsler Adult Intelligence Scale. New York: The Psychological Corporation.

Wechsler, D. (1967). Manual for the Wechsler Preschool and Primary Scale of Intelligence. New York: The Psychological Corporation.

Wechsler, D. (1974). WISC-R Manual: Wechsler Intelligence Scale for Children - Revised. New York: The Psychological Corporation.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children - Third Edition Manual. San Antonio: 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. [Monograph series: Advances in psychoeducational assessment: Wechsler Intelligence Scale for Children:Third Edition]. Journal of Psychoeducational Assessment, 34-42.

Wherry, R. J., & Wherry, R. J., Jr. (1969). WHEWH program. In R. J. Wherry (Ed.), Psychology department computer programs. Columbus, OH: The Ohio State University, Department of Psychology.

Williams, R. L. (1970). Danger: Testing and dehumanizing Black children. The Clinical Child Psychology Newsletter, 9, 5-6.

Williams, R. L. (1971). Danger: Testing and dehumanizing Black children. The School Psychologist, 25, 11-13.

Wilson, J. J. (1987). The truly disadvantaged: The inner city, the underclass, and public policy. Chicago: University of Chicago Press.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. Journal of Psychoeducational Assessment, 8, 231-258.

Woodcock, R. W. & Johnson, M. B. (1989). Woodcock-Johnson Psycho-Educational Battery -- Revised. Chicago: Riverside.

Wornhoff, S. A. (1997). The changing role of response speed on the three editions of the Wechsler Intelligence Scale for Children. Unpublished doctoral dissertation, Indiana University, Bloomington.

Zachary, R. A. (1990). Wechsler's Intelligence Scales: Theoretical and practical considerations. Journal of Psychoeducational Assessment, 8, 276-289.

Zarske, J. A., Moore, C. L., & Peterson, J. D. (1981). WISC-R factor structures for diagnosed learning disabled Navajo and Papago children. Psychology in the Schools, 18, 402-407.

Zhu, J., Woodell, N. M., & Kreiman, C. L. (1997, August). Three year reevaluation stability of the WISC-III: A learning disabled sample. Paper presented at the 105th Annual Convention of the American Psychological Association, Chicago, IL.

APPENDIX A
THE WISC-III

The WISC-III

Characteristics of the Standardization Sample of the WISC-III

The primary goal in the development of the WISC-III was to update the norms of the WISC-R, which had been published in 1974. Many practitioners believed this was long overdue. The importance of maintaining current norms had been established from extensive research. In her summary of this research, Anastasi (1988) noted that when conditions in a culture improve over time “including increasing literacy, higher educational levels, and other cultural changes,” they are accompanied by significant rise in intellectual performance (p. 353).

Flynn (1994) concluded from his extensive study of this phenomenon, which he called “upward drift,” that IQ scores rose 1/3 to 1/2 point per year on average in the United States as well as in Europe. The effect of this “upward drift” on IQ scores according to Wechsler (1991) is that they can give “a progressively deceptive picture of a child’s abilities relative to others in the same age group” (Wechsler, 1991, p.4). Wechsler suggested problems posed by this tendency for a child’s score to be higher with outdated norms than current norms could be greater for children who earn scores well above or well below average. Because the mean IQ scores of children with learning disabilities reported in some studies tended to be below average (MacMillan et al., 1996) the upward drift phenomenon might impact more significantly on some members of that clinical group. Inflated IQ scores when contrasted with achievement scores in discrepancy analysis could make it possible for some children to be classified learning disabled where a true score discrepancy does not exist, for example.

The WISC-III was normed based upon an analysis of the data gathered by the United States Bureau of Census in 1988. Stratification in the norming sample for the WISC-III was accomplished based upon these data “along the following variables: age; gender; race/ethnicity; geographic region; and parents’ education (parent refers to parent(s) or guardians(s))” (Wechsler, 1991, p. 20). The stratification sample included a total of 2,200 cases. One hundred male and 100 female children were included at each of 11 age groups from ages 6-16. The mean age for each age group was reported as “the sixth month (e.g., 6 years, 6 months; 7 years, 6 months; 8 years, 6 months, etc).” (Wechsler, 1991, p. 20).

Geographical distribution in the United States was reported using the categories specified in the United States Census reports: Northeast; South; North Central; West. “Children were selected for the normative group in accordance with the proportions of children living in each region” (Wechsler, 1991, p. 20). Community size was considered as well. The percentages of the WISC-III standardization sample from metropolitan areas with populations of over 1,000,000 was 36.7 versus 43.5 in the overall United States population based upon the 1988 United States Bureau of Census data (Wechsler, 1991).

Representation by ethnicity for each age group (i.e., Whites, African Americans, Hispanics and Other [including Native American, Eskimo, Aleut, Asian, and Pacific Islander]) was developed “based on the race/ethnic group proportions of children aged 6-16 in the U.S. population according to the 1988 census survey” (Wechsler, 1991, p. 20). Parent reports were utilized as the identifiers for racial categorization of subjects.

The socioeconomic status (SES) of the sample was categorized using level of parent education. SES is especially important in the development of norms for tests of intelligence because differences in SES are associated with large differences in scores on intelligence tests (Kamphaus,1993). The education level of parents for the WISC-III standardization sample was evaluated and reported in “a matrix of five parent education levels by child/race ethnicity for each combination of age, group, gender, and geographic region” (Wechsler, 1991, p. 21). The five education categories were: 8th grade or less; 9th - 11th grade; high school graduate or equivalent; 1 - 3 years of college or technical school; 4 or more years of college. The face validity of parent reports was accepted as proof of parent education level. In instances where children lived with two parents the average of their education level was used.

Granier and O'Donnell (1991) evaluated the relationship between IQ and level of parent education specifically for this sample. They used a portion of the total standardization sample (i.e., 1, 194 children, ages 6-16). Results of their study showed that children's mean IQ varied linearly in relation to parent education. Children whose parents had completed college had the highest IQ's (mean IQ = 106.01). This was followed next by children whose parents had some college (mean IQ = 100.82). Children whose parents had a high school diploma (mean IQ = 97.72) earned slightly lower scores. Children whose parents attended grades 9 through 11 (mean IQ = 92.10) and finally children whose parents had less than a 9th grade education (mean IQ = 86.38) earned the lowest scores. Their findings were similar to those obtained with the WISC-R sample even though parent occupation rather than

parent education had been utilized as the stratification variable to discriminate socioeconomic status.

The stratified racial distribution by geographic region for the standardization sample of the WISC-III reported in the manual revealed that the majority of white and African American samples came from the North Central and South regions (White = 61.2%; African Americans = 81.2%). The majority of Hispanic and other samples were taken from the South and West (Hispanic = 84.5%; other = 66.6%).

The total of 2,200 students included in the stratification sample were drawn from both public and private schools. The percentage from public versus private schools is not reported in any category described previously herein in developing the stratification matrix.

It is noted that 7% of the total standardization sample consisted of children classified as learning disabled, speech/language impaired, emotionally disturbed, physically impaired, or in Chapter I programs (Wechsler, 1991, p. 22). This would reflect a total of about 154 of the 2,200 cases. No information is available regarding age, gender, race, parental education, geographical considerations, or number of cases in each category. All children included in this clinical sample were drawn from mainstream classes (personal communication with Dr. Charles Wilkins, Research Analyst for The Psychological Corporation, 2/15/99).

There is consensus that the technical quality of standardization of the WISC-III is viewed favorably. Sattler (1992) compared the procedures used in WISC-R norming to those employed with the WISC-III and concluded the sampling procedures for the WISC-III were “notably superior” (p. 1034).

Preliminary data in the WISC-III manual suggests that all children, regardless of ability level and clinical status, score 5 points lower on the WISC-III than they do on the WISC-R (Wechsler, 1991). This includes not only non-handicapped children and gifted children but also children with handicapping conditions including learning disabled and attention deficit. Mentally retarded students scored 9 points lower. Score variation is congruent overall with Flynn's (1997) assertion of population IQ gains of 1/3 to 1/2 point each year.

Content Revision in the WISC-III

A goal in the development of the WISC-III was to maintain “the essential structure and content of the WISC-R” (Wechsler, 1991, p. 11). All original subtests were retained. Extensive studies of WISC-R item bias were conducted regarding gender, ethnic, and regional bias. The review panel included minority experts. In order to identify racial/ethnic bias more clearly in test items, item analysis procedures were utilized to evaluate the results of the performance of 400 ethnic minority children. Items were altered, rejected or replaced only when it was shown that they had lost their cultural meaning or did not discriminate fairly. About 73% of the items were in fact retained from the WISC-R according to Aurelio Prifitera, Project Director in The Psychological Corporation for the WISC-III (personal communication, October 5, 1992).

Concern about provision of a more accurate downward and upward measurement on some subtests resulted in the addition of some items. The following subtests on the WISC-III now contain additional items: Similarities; Arithmetic; Comprehension; Digit Span; Picture Completion; Coding; Object Assembly; Mazes;

Picture Arrangement; and Block Design. O'Donnell, Granier, and Dersh (1991) investigated the effect of handedness on WISC-III design modifications and alternative administration procedures for left-handed children. They concluded the “modifications and alternative administration procedures for left-handed children has not resulted in WISC-III differential Coding performance” (O'Donnell et al., 1991, p. 7).

The addition of color on Picture Completion, Picture Arrangement, and selected Object Assembly items was accomplished to enhance motivation of children. The order of presentation of the initial subtests was altered to address this issue as well. The first subtest presented in the WISC-III is Picture Completion which is considered to be less intimidating to younger children and children with learning problems.

A single new optional subtest, Symbol Search, was developed as an outgrowth of efforts to study complex aspects of the third factor involving controlled attention research and research on memory scanning abilities (Wechsler, 1991, p. 12). Like Coding, this subtest contained two developmental levels, with each level containing 45 items. Although the subtest was developed to buttress the third factor, research including factor analytic studies, suggested it joined with Coding and created a new fourth factor, labeled Processing Speed (Wechsler, 1991).

The addition of Symbol Search, which has a speed of response component, and expansion of Picture Arrangement and Object Assembly to include more timed items with bonus points raised a question about whether the fundamental structure of test had been altered with respect to emphasis on speed of response. Kaufman

(1993), in his brief analysis of speed of response on the WISC-III, reported that there was greater emphasis on this construct. This important question was addressed more in greater depth subsequently by Wornhoff (1997).

Wornhoff (1997) conducted a more comprehensive investigation of the role of response speed across all three editions of the WISC. He reported the role of response speed increased most noticeably from the WISC-R to the WISC-III, with the number of possible bonus points being assigned to speed increasing from 51 on the WISC-R to 86 on the WISC-III. In his investigation, he found the strongest impact of increased bonus points occurred upon children 12 years of age and older who failed to earn bonus points on Picture Arrangement, Block Design, and Object Assembly.

The impact of Picture Arrangement bonus points was by far the most powerful determinant of changes in scores related to response speed for these older children. The difference between the difficulty involved in earning bonus points from the WISC-R to the WISC-III regarding Block Design and Object Assembly subtests was not significant for children in any age group according to Wornhoff's research. In the 12-16 year old age groups, WISC-III Picture Arrangement subtest bonus points accounted for 40% of the total points that could be earned on that subtest compared to 24% on the WISC-R Picture Arrangement subtest. Wornhoff calculated the impact of these bonus points upon the WISC-III FSIQ and concluded this subtest, "will typically account for only 1 to 2 IQ points in the overall Full Scale IQ" (Wornhoff, 1997, p. 89). His findings therefore strongly dispute Kaufman's conclusion that the WISC-III is significantly more dependent upon response speed than the WISC-R.

APPENDIX B
STUDIES OF PATTERNS OF ABILITIES OF SAMPLES USING THE WISC-III
SUBTEST SCORES

Studies of Patterns of Abilities of Samples Using the WISC-III Subtest Scores

In the most recent analysis, T. J. Ward, S. B. Ward, Glutting, and Hyatt (1999) examined ability and achievement profiles generated from WISC-III and WIAT scores of 201 children identified as having learning disabilities. Initially, results of a hierarchical cluster analysis revealed five distinct clusters reported to be similar to those found in the previous research of Glutting et al. (1994). Two clusters were compatible with criteria from a discrepancy model (i.e., ability/writing and VIQ/PIQ as well as ability, reading and writing). One group resembled the low ability/ low achievement cluster from the Glutting et al. (1994) study. The final group resembled the below average ability/below average achievement group of “slow learners” from the Glutting et al. (1994) study. These four groups cumulatively demonstrated profiles compatible with the results drawn from the WISC-III/WIAT linking sample without disabilities in 70% of the LD sample.

In the second cluster analysis the scores of the children whose profiles were distinct from the six core profiles of the WISC-III/WIAT linking sample (Glutting et al., 1994) were analyzed. Two subtypes emerged. The first group “resembled slow learners with consistently below average ability and achievement” (Ward, T. J. et al., 1999, p. 639-640). This pattern was characteristic of Glutting et al.’s (1994) profile type 6. Although performance of the group on oral expression was significantly lower (but not low enough to constitute a significant discrepancy from ability) than Glutting et al.’s (1994) profile type 6, the similarities extended into the fact this group too, like Glutting et al.’s, was predominantly of minority status (79%). In this group 71% were male and the median age of the group was 12.9 years, reflecting a cluster

of somewhat older, male, minority children with learning disabilities (T. J. Ward et al., 1999).

The second group, predominantly Caucasian (81%), displayed significant ability, achievement discrepancies in reading and writing. The magnitude of difference between ability, achievement scores (i.e., more severe fluctuations was what distinguished the scores of this group from those of the core subtypes in the Glutting et al. (1994) study. This group was composed of equal numbers of males and females. The average age of these groups was 10.5 years.

Konold, Glutting, McDermott, Kush, and Watkins (1999) investigated the variations in score base rates of the 10 core subtests for the WISC-III standardization sample using nonlinear multivariate Q methodology “simultaneously according to the level and shape of their subtest scores” (pp. 31-32). Konold et al., used this analysis to construct a normative taxonomy of the most common subtest profiles for the WISC-III standardization sample. Eight core profile types were identified, distinguished primarily by differences in general ability level. Konold et al., stress the finding that the profiles are not flat within the eight core profile types.

Profile 1, “High Ability,” reflects the highest FSIQ mean (126.20). A higher proportion of Caucasian children were present in this group. No African American children were present and fewer Hispanics than anticipated are reported. Profile 2, “Above Average Ability,” (mean FSIQ = 113.90) also included a higher proportion of Caucasians than is found in the general child population and less than a quarter of the proportion of African Americans expected. Fewer Hispanics were present than anticipated. About 65% of the children in this category were girls. Profile 3, “Above

Average Ability,” was distinguished both by FSIQ (mean = 108.50) and a VIQ/PIQ discrepancy favoring VIQ (mean discrepancy = 8.5 points). Again, a higher proportion of Caucasians were present than anticipated but fewer African Americans and Hispanics. About 65.8% of this group were boys. This profile showed an age effect as well with more 13-16 year olds and fewer 6-8 year olds than anticipated.

Two average ability profiles emerged. Profile 4, “Average Ability and PIQ>VIQ,” (mean FSIQ = 102.60) exhibited the greatest disproportion of VIQ/PIQ differences. A larger number of PIQ>VIQ discrepancies emerged (mean difference = 11.2 points). More boys than girls were present in this core group (65.6%). Profile 5, “Average Ability and VIQ>PIQ,” (mean FSIQ = 99.10) reflected a higher occurrence of unusual VIQ>PIQ discrepancies (mean difference = 6.2 points). More girls and slightly more Caucasian children were present in this group than expected.

Two below average groups emerged. Profile 6, “Below Average Ability and PIQ>VIQ,” (mean FSIQ = 89.30) has a prevalence rate of 12.9%. It reflected the second largest disproportion of VIQ/PIQ differences, with more PIQ>VIQ discrepancies appearing here (mean difference = 10.7 points). More Hispanics and less Caucasians were present. Profile 7, “Below Average Ability,” (mean FSIQ = 87.60) has a prevalence rate of 14%. The frequency of PIQ>VIQ discrepancies was lower than expected. More boys (57.3%) than girls were represented in this profile. Over twice as many African Americans appeared in this group as are expected. Fewer Caucasians appeared than expected. More children in special education programs appeared in this profile than were expected.

The final profile, Profile 8, “Low Ability” (mean FSIQ = 73.10) has a prevalence rate of 8.4%. Over two and one-half times the number of African Americans appeared in this category as would have been expected. Fewer Caucasians and Hispanics were present than were expected. More children from the South appear in this group than were expected and the percentage of children from the Northeast was below expectancy. More children in special education programs appeared than were expected.

Glutting et al., (1997) utilized nonlinear-multivariate methodology to identify the most representative subtest patterns for the standardization sample of the WISC-III using the scores from the 13 subtests. These scores were compared one at a time to the mean score regarding each child’s personal performance on the WISC-III. The normative profile taxonomy simultaneously grouped children in the standardization sample according to the level (above average, average, below average ability) and the shape (variations in pattern across subtest scores) of their subtest scores across all eleven age levels utilizing cluster analysis. The portion of the standardization sample classified as learning disabled, emotionally disturbed, physically impaired, and speech/language impaired were collapsed into one category for purposes of statistical comparison.

Nine core profiles were identified in the statistical analysis. These core types displayed the greatest differences with respect to *g*. Additionally four of the profile types reflect differences between the VIQ and PIQ (i.e., profile types 2, 3, 7, and 8). The prevalence of severe FSIQ/FDI discrepancies in types 5 and 9 and the prevalence of severe FSIQ/PSI discrepancies in types 1, 2, 3, 7, and 9 are noted as well.

Three of the profiles involved above average FSIQ. The first, Type 1, labeled High Ability and Depressed Processing Speed is associated with an average FSIQ of 125.6. It occurred in 9.4% of the standardization sample. The proportion of African American children in this group was very small, occurring at less than 1/100th of the level anticipated from the 1988 U.S. census data. The second type, Type 2, labeled Above Average Ability with VIQ>PIQ and Depressed Processing Speed was associated with an average FSIQ of 114.1. It occurred in 13.3 % of the standardization sample. Fewer African American children were reported than anticipated in this category as well. The third type, Type 3, labeled Above Average Ability with PIQ>VIQ and Elevated Processing Speed was associated with an average FSIQ of 108.60. It occurred in 14.1 % of the standardization sample. The Type 3 group included the portion of African American children at their expectancy level.

Two profiles were characterized as having average FSIQ scores. Type 4, “Average Ability,” is associated simply with an average FSIQ of 101.60. It occurred in 13.2 % of the standardization sample. It included only half the number of African American children that were expected. Type 5, “Average Ability and Elevated Freedom from Distractibility,” is associated with an average FSIQ of 99.20. It occurred in 9.5 % of the standardization sample. The Type 5 group included the portion of the African American sample at their expectancy level.

Four profiles were characterized as having below average FSIQ scores. Type 6, labeled simply Slightly Below Average Ability, is associated with an average FSIQ of 92.40. It occurred in 9.9 % of the standardization sample. No differences were

reported for the portion of the African American sample that was represented in this category. Type 7, labeled Slightly Below Average Ability with PIQ>VIQ and Elevated Processing Speed, is associated with an average FSIQ of 91.30. It occurred in 12.1 % of the standardization sample. It reflected no differences in the portion of the sample of African American children that was represented in this category. Type 8, labeled Below Average Ability and VSIQ>PSIQ is associated with an average FSIQ of 86.0. It occurred in 9.7 % of the standardization sample. It included over twice as many African American children than would have been expected. The last core profile, Type 9, labeled Low Ability with Elevated Freedom from Distractibility and Elevated Processing Speed, is associated with an average FSIQ of 73.60. It occurred in 8 % of the standardization sample. “The frequency of African Americans was over two and one-half times higher than the national average” (Glutting et al., 1997, p. 33). Additionally, Glutting et al. (1997), report “A leaning is present for an overrepresentation in special education programs” (p. 33).

Donders (1996) examined the Index scores in the standardization sample of the WISC-III using SAS cluster analysis and Fastclus procedures to identify core profile subtypes. Five cluster subtypes emerged, three of which were differentiated almost exclusively by level of performance and two which were differentiated by patterns of performance. Three clusters, Cluster 4, Cluster 3, and Cluster 2 are differentiated by level of performance across all four Indexes. The levels of performance are characterized as above average, average, and below average. Cluster 3 has four Index scores more than one standard deviation above average. Cluster 4 has average scores on all Indexes. Cluster 2 has below-average scores on the four

Indexes. Donders reported that differences between the highest and lowest average Index score were less than 7 points in these three clusters.

Cluster 1 and Cluster 5, by contrast are characterized, in part, by differentiated performance on the Processing Speed factor. Cluster 1 has average VC, PO, and FD scores and a PS score that is more than 13 points higher than the other three clusters. Cluster 5, by contrast, has PS scores 9 - 12 points lower on average than other factor scores that primarily are in the average range.

Donders (1996) reported that age did not emerge as a significant variable in the cluster solutions. Mean level of parental education did emerge as a statistically significant variable, however. More than a third of the parents in Cluster 3 had 16 or more years of education and the mean factor scores of the children in this cluster consistently were the highest for the standardization sample. More than 40 % of the parents of Cluster 2 children had less than 12 years of education and the mean factor scores of the children in this cluster were the lowest of the five cluster subtypes.

Donders (1996) concluded the pattern of WISC-III Index scores is not necessarily flat for many children in the standardization sample. This finding has implications for clinicians who examine the significance of discrepancy data in Index scores to assist them in making decisions about whether a child has a learning disability. The pattern of performance that was reported by Donders in relation to the standardization sample suggests that significant variations, in and of themselves, cannot be considered indicators of disabilities.

Glutting, McDermott, Prifitera, and McGrath (1994) analyzed multivariate IQ - achievement discrepancies in a sample of 824 children from the WISC-III and

WIAT (Wechsler Individual Achievement Test) linking sample of 824 children using multivariate cluster analysis (Q methodology). Six distinct profiles were obtained that were distinguished largely by overall level of performance. The first profile was characterized by the highest average FSIQ, the highest level of achievement and $VIQ > PIQ$. The second profile was characterized by above average FSIQ, slightly above average achievement, and $PIQ > VIQ$. The third profile reflected average FSIQ and underachievement in writing. Average FSIQ and overachievement in reading, mathematics, language, and writing characterized the fourth profile. Below average FSIQ and below average achievement characterized the fifth profile. Low ability (i.e., low FSIQ, underachievement in reading, mathematics, and writing, as well as $PIQ > VIQ$) characterized Profile 6.

Glutting et al. (1994) note that “approximately one-half of the most common profile types for the WISC-III and WIAT are defined by disproportions in the number of children showing unusual univariate FSIQ - achievement discrepancies ... children from the general population normally show score strengths and weaknesses” (p. 629). The frequency of occurrence in this subset of the normative population is noteworthy. It suggests discrepancy analysis using the WISC-III and achievement measures should take into account the magnitude of natural variation that exists in the scores of the normative population. It also underscores the importance of using co-normed IQ/achievement instruments that can present precise statistical evidence of the prevalence of this variation in the normal population.

Some fluctuations in patterns of performance were reported. Congruent with the findings of the 1997 Glutting et al., study, more African American children were

found than expected in Types 5 and 6 in the co-norming study. The Type 5 profile accounted for 17.6 % of the sample and the Type 6 profile accounted for 8.5 % (Glutting et al., 1994). Of special interest to researchers studying patterns of performance of learning disabled children, is the similarity between the Type 6 profile and the profile of children with nonverbal learning disabilities (NLD), with two notable exceptions. Those exceptions are that NLD children generally test with higher ability scores and higher reading scores than this subset demonstrate.

The Type 6 profile is the only profile to show an age effect, a gender effect, and a race effect as well. A greater number of older children (ages 8, 9, and 10) were found in this group. Nearly 70 % of the children were female. The frequency of African American and Hispanic children was twice the rate of the general population (Glutting et al., 1994). Glutting et al., reported that there were no systematic differences in the education level of the parents of the children in the Type 6 profile. Because the data involved small absolute numbers of African American children the findings should be viewed with caution. It is interesting to note, however, that with regard to ability, there is congruence with the data analyzed from the full standardization sample in 1999 by Konold et al., and 1997 by Glutting et al. That sample was larger and carefully stratified to reflect the 1988 U.S. census. All four studies underscore the importance of recognizing the prevalence of score discrepancies in the normal population of school-age children.

S. B. Ward, T. J. Ward, Hatt, Young and Mollner (1995) investigated the incidence and utility of three WISC-III subtest patterns in the differential diagnosis of exceptionalities in a sample of 719 children referred for assessment in a southeastern

urban center. The sample included 382 children with learning disabilities, 88 children with emotional disabilities, 42 children with mental retardation, and 172 children who were determined to be ineligible for special education services. The ethnicity of the sample was reported to include 69.3% Caucasian children, 26.9% African American children, 2.5% Hispanic children, .9% Asian children, and .4% Native American children.

Three profiles associated with exceptionality were studied. The first profile, the ACID profile, consisting of scaled scores on Arithmetic, Coding, Information, and Digit Span, had been reported in WISC-R studies to be associated with depressed scores for children with learning disabilities (Reynolds & Kaufman, 1990; Sandoval, 1984). Additionally, it was reported for two samples of children with learning disabilities with the WISC-III (Prifitera & Dersh, 1993; Wechsler, 1991). The second profile, the ACIDS profile, consisting of scaled scores from Arithmetic, Coding, Information, Digit Span, and Symbol Search, was investigated by Prifitera and Dersh (1993). It was found to be more prevalent in their sample of children with attention deficit disorders than children with learning disabilities.

The third profile, the SCAD profile, consists of scaled scores on Symbol Search, Coding, Arithmetic, and Digit Span. Prifitera and Dersh (1993) investigated it in conjunction with the Perceptual Organization Index of the WISC-III. Prifitera and Dersh found the magnitude of PO/SCAD differences to be greater in their samples of children with learning disabilities and attention deficit disorders than the WISC-III standardization sample.

Ward et al. (1995) found that the ACID and ACIDS profiles occurred infrequently in their sample, 3.4% and 1.2% respectively. The SCAD profile occurred for 15.4 % of their total sample for whom Symbol Search scores were reported and 19.6% of the subsample with learning disabilities for whom Symbol Search scores were reported. This subtest had not been administered to every child in the sample. Additionally, children with attention deficit disorders were not listed as a separate category in their sample. Ward et al., concluded at best the ACID and SCAD profiles served as a confirmatory variable for very few children. The ACIDS profile demonstrated no evidence of utility.

APPENDIX C
EQS NOTATIONS AND PATH DIAGRAMMS

EQS Notations and Path Diagrams

Using EQS notation, all variables are considered to fall into two categories. They included measured (observed) variables, designated as V's, that are drawn from the actual data. In this study that data was the WISC-III subtest scores. All other variables represent the structural network or the unmeasured variables (Byrne, 1994b). Three kinds of unmeasured variables are given EQS notations. The latent constructs or factors are designated as F. The residuals associated with the prediction of each factor are designated as D's for disturbance terms. The residuals associated with the measurement of every observed variable are designated as E's for error. The E's and D's are given numbers corresponding to the V's and F's with which they are associated.

The schematic representation of the model is called a path diagram. The path diagram visually portrays the relationships assumed to exist among the variables. Certain conventions govern path model construction and representation. Measured variables are the observable data, usually obtained from scores on measures. Boxes represent them. Each box is labeled V. The measured variables, V's, "function as indicators of their respective underlying latent factors" (Byrne, 1994b, p. 8). The error term associated with each V reflects the variance contributed by measurement error and the specificity of the measured variable that is not accounted for by the latent variable. It is assigned a number corresponding to its V.

The unmeasured or latent variables, which are called factors in CSM, are shown in ellipses or circles. The disturbance term, or residual, associated with each factor is typically enclosed in an ellipse as well. This small ellipse is located close to

its larger factor ellipse. It bears the same number as the number assigned to the factor with which it is associated.

Hypothesized processes assumed to involve the entire system of variables are denoted with symbols in the forms of unidirectional arrows, representing structural regression coefficients (Byrne, 1994b). These structural regression coefficients “indicate the impact of one variable on another” (Byrne, 1994b, p. 8). The path diagram pictorially portrays the direction and strength of each variable, as indicated by its path arrow and path coefficient (Jensen, 1998). The head of the unidirectional arrow flows always from the causal variable (latent variable) either to an observed variable, indicating this observed variable was “caused” (Byrne, 1994b, p. 9) by the latent variable, or to another latent variable which the arrow implies it caused. Byrne characterizes the arrows associated with residuals as being “sourceless one-way arrows” (Byrne, 1994b, p. 9). Additionally, curved bi-directional arrows are used to represent both covariances and correlations between pairs of variables. These can occur only between latent independent variables.

In the EQS SEM program developed by Bentler and Weeks (1980), the causal (latent) variables are characterized as being the independent variables. The observed variables represent dependent variables. A dependent variable “can be expressed as a structural regression function” (Byrne, 1994b, p. 10) of an independent variable, as reflected in an equation. EQS provides the mathematical representation of the schematic presentation of the model, with the addition of some parameters not specified in the model that are critical to the mathematical representation of the relationships of the variables (Byrne, 1994b).

Path diagrams were constructed in the present study for the partially restricted factor analysis of the CN, AAN, and CD samples for the two-factor, three-factor, and four-factor models. These diagrams were presented in Chapter IV. In these path model diagrams, rectangular boxes represented measured variables (i.e., WISC-III subtests). The actual subtest names rather than V labels were applied in each diagram to illustrate the components of each model more clearly. The values of the structural regression coefficients for the measured variables were indicated for every path drawn from the latent variables (i.e., factors). The square of the path coefficient of the error term for each observed variable, which reflects the combined proportion of error variance not explained by the latent variable, was reported as well. Factor covariances were reported and indicated beside the curved bi-directional arrows that demarcate covariance.

APPENDIX D

GOODNESS-OF-FIT INDEXES REPORTED IN EQS

Goodness-of-Fit Indexes Reported in EQS

Stand-alone or absolute indices directly assess “how well an a priori model reproduces the sample data” (Hu & Bentler, 1998, p. 426). They are comparable to an R squared, which indicates the proportion of variance accounted for by the a priori model. These include LISREL’s Goodness-of-Fit index (GFI), Adjusted Goodness-of-Fit index (AGFI), and the McDonald Fit index (MFI). Additionally they include the Unstandardized Root-Mean-Square Residual (RMR).

Joreskog and Sorbom developed the GFI in 1984 (Hu & Bentler, 1998). According to Joreskog and Sorbom, it is purported to measure the amount of the variances and covariances accounted for jointly by the model. The GFI in SEM is the equivalent of the R squared in multiple regression. Joreskog and Sorbom assert the GFI is independent of sample size but the work of Marsh et al. (1988) disputes this sharply. The GFI does have a maximum value of unity and it can be negative (Hu & Bentler).

Joreskog and Sorbom also developed the AGFI in 1984. It “assesses the amount of variation/covariation in the sample covariance matrix”(Konold et al., 1997, p. 128) that the model predicts. It estimates the variance accounted for in the population. The AGFI uses mean squares in place of sums of squares, thus incorporating a penalty function for additional parameters (Marsh et al., 1988). Although Joreskog and Sorbom assert the AGFI is independent of sample size, the work of Marsh et al. (1988) dispute this sharply as well. The AGFI has a maximum value of unity and it can be negative (Hu & Bentler, 1998).

The MFI (McDonald's Centrality index) was developed by McDonald in 1989 (Hu & Bentler, 1998). It is "an absolute goodness-of-fit index (as opposed to indices that are relative to a null or other alternative model) which by its form is not systematically dependent on sample size" (McDonald & Marsh, 1990). Its complexity statistically involves initially defining "a population-fit index parameter" (Hu & Bentler, p.427). Estimators from this parameter are then used to define the index reflecting fit of the sample (Hu & Bentler). It typically has a range from zero to unity but the upper bound may exceed 1 (Hu & Bentler) in the case where sampling error has a marked effect (McDonald & Marsh).

Joreskog and Sorbom developed the RMSR in 1981. They defined it as "the square root of the mean of squared residuals" (Joreskog & Sorbom, 1981, p. I.41) between the sample matrix and the fitted population covariance matrix. The RMSR can be viewed as "a measure of the degree of reproduction of the covariance matrix from the model estimates" (Wechsler, 1991, p. 194). "RMSR represents the absolute value of the average fitted residuals for a given CFA model" (Bryant & Yarnold, 1995, p. 132). When correlation matrices serve as the basis for the sample matrix and fitted population matrix, the values of RMSR are bounded by zero and unity. When covariance matrices are used, RMSR has a lower bound of zero but it does not have an upper bound (Marsh et al., 1988). Therefore RMSR is interpretable only "in relation to the size of the variances and covariances of the measured variables and cannot be compared across applications based on different variables" (Marsh et al., p. 392). When the residuals yields negative values they indicate the model is

overpredicting the covariance matrix. When they yield positive values they indicate the model is underpredicting the covariance matrix (Bollen, 1989).

Incremental or comparative fit indices measure improvement in fit. A target model is compared to a more restricted, nested baseline model, which typically is a null model. In the null model, “all the observed variables are allowed to have variances but are uncorrelated with each other” (Hu & Bentler, 1998, p. 426).

Examples of incremental fit indices included the Bentler-Bonett Normed Fit index (NFI), the Tucker Lewis index (TLI) or Bentler-Bonett Nonnormed Fit index (NNFI), the Bollen Incremental Fit index (IFI), and the comparative fit index (CFI).

The NFI, developed by Bentler and Bonett in 1980, has enjoyed such popularity that it is characterized by Byrne (1994b) as being “the practical criterion of choice” (p.55). In certain contexts the NFI is considered to reflect “the proportion of total information” (Mulaik et al., 1989, p. 432) about associations between variables for which a model accounts. The range of the NFI is zero to unity. The NFI appears to be moderately sensitive to complex model misspecification but not sensitive to simple model misspecification (Hu & Bentler, 1998). Marsh et al. (1988) report that for small samples (fewer than 200) the NFI is problematic. Bentler (1990a) also reports the NFI “is downward biased in small samples” (p. 166).

Tucker and Lewis developed the TLI in 1973. According to Tucker and Lewis, it was developed as a reliability coefficient “to indicate quality of representation of interrelations among attributes in a battery by a factor analytic model having a limited number of common factors” (Tucker & Lewis, 1973, p. 9). It was developed initially to be used in ML exploratory factor analysis. Bentler and

Bonett (1980) built the Nonnormed Fit index (NNFI) based upon the TLI, incorporating a degree of freedom adjustment in the index to improve its performance near 1.0 (Bentler, 1990b). In confirmatory factor analyses, the TLI is computed “by comparing the likelihood ratio chi-square and df from a theoretically derived model to a baseline model” (Kamphaus et al., 1994, p. 177). The TLI is a nonnormed index that can fall outside the bounds of zero to unity. Bentler and Bonett (1980) recommend the TLI as being useful in comparing the fit of a model across samples that are characterized as having unequal sizes.

Marsh et al. (1988) report the TLI to be relatively independent of sample size. Hu & Bentler (1998), however, caution the nonnormed TLI is less preferable in sample sizes under 250. Bentler (1990a) recommends the Bentler-Bonett Nonnormed Fit index (NNFI) over the NFI in situations wherein small samples are used. Bentler (1990a) reports the NNFI has a large sampling variance. Hu and Bentler additionally found the TLI to be on of the indices most sensitive to models that had misspecified factor loadings. Additionally they note it also does incorporate a penalty function for nonparsimonious models. The TLI was reported as one of the multiple indices of model-data fit reported in confirmatory factor analyses of the WISC-III standardization sample data (Wechsler, 1991).

Bollen provided the Bollen Incremental fit index (IFI) in 1989. It too is nonnormed and compensates for the effect of model complexity (Hu & Bentler, 1998). It is among the indices more sensitive to models with misspecified factor loadings (Hu & Bentler). Marsh et al. (1988) report that it is substantially affected by sample size, however, especially when the sample size is less than 200.

The comparative fit index was developed by Bentler (1990a) as a revision of the NFI to take into account small samples size (Byrne, 1994b). It has many advantages over the NFI and NNFI in small samples according to Bentler. He reports it eliminates small sample bias, has a smaller sampling variance than the NNFI. It “estimates the relative difference in noncentrality of interest” (Bentler, 1990b, p. 245). It remains in the range of zero to unity. Like the NFI, it is derived from a comparison of a hypothesized model with a null model, providing “a measure of complete covariation in the data” (Byrne, 1994b, p. 55) without requiring knowledge of sources of misspecification. It is among the indices more sensitive to models with misspecified factor loadings (Hu & Bentler, 1998). It does not include a penalty function for nonparsimonious models, however (Hu & Bentler). Byrne accepts a value greater than .90 with respect to the CFI in examining its goodness-of-fit.

APPENDIX E

SELECTED FIT STATISTICS FOR PARALLEL PARTIALLY RESTRICTED
FACTOR ANALYSES WITH CONSTRAINTS FOR AGE COLLAPSING OF
AFRICAN AMERICAN CHILDREN WITH LEARNING DISABILITIES IN
TWO, THREE, AND FOUR-FACTOR MODELS

Selected Fit Statistics for Parallel Partially Restricted Factor Analyses with
Constraints for Age Collapsing of African American Children with Learning
Disabilities in Two, Three and Four-Factor Models

<u>Age Range</u>	<u>Two-Factor Model</u>		<u>Three-Factor Model</u>		<u>Four-Factor Model</u>	
	(n = 976)		(n = 646)		(n = 172)	
	<u>TLI^a</u>	<u>CFI^a</u>	<u>TLI^a</u>	<u>CFI^a</u>	<u>TLI^a</u>	<u>CFI^a</u>
All (6-16)	.92	.94	.91	.94	.90	.93
11 and 12	.97	.97	.97	.97	data insufficient ^b	
11-12 and 10	.94	.94	.91	.92	.80	.82
10-12 and 13	.92	.93	.90	.91	data insufficient ^b	
10-13 and 9	.89	.89	.74	.76	.60	.64
9-13 and 14	.89	.90	.87	.88	data insufficient ^b	
9-14 and 8	.79	.79	.68	.71	.81	.83
8-14 and 15	.92	.92	.91	.92	.83	.85
8-15 and 7	.70	.71	.58	.62	.87	.88
7-15 and 16	.92	.93	Data insufficient ^b		data insufficient ^b	
7-16 and 6	.89	.90	.88	.89	data insufficient ^b	

Note. Sample sizes were inconsistent across the age ranges. ^aTLI = Tucker-Lewis index; CFI = comparative fit index. ^bSmall numbers for certain age groups result in insufficient data to iterate even one time in parallel partially restricted factor analyses.

APPENDIX F
FIT INDICES FOR FULLY RESTRICTED FACTOR ANALYSES FOR THE
TWO, THREE, AND FOUR-FACTOR MODEL

Fit Indices for Fully Restricted Factor Analyses for the Two, Three,
and Four-Factor Model

Table F1

Model Evaluation Statistics for the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Two-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AAN</u>	<u>CD</u>	<u>AALD</u>
χ^2	88.48	166.87	263.86
df	43	43	43
BBN	.94	.88	.88
TLI	.96	.90	.90
CFI	.97	.91	.90
IFI	.97	.91	.90
MFI	.94	.74	.89
GFI	.95	.85	.95
AGFI	.94	.80	.93
RMR	.57	3.89	1.47
SRMSR	.06	.27	.16
RMSEA	.06	.12	.07
90% Confidence Interval of RMSEA	.04, .07	.10, .14	.06, .08

Note. AAN = WISC-III African American Sample Without Disabilities; CD = WISC-III Caucasian Validity Sample With Disabilities; AALD = African American Sample With Learning Disabilities; BBN = Bentler-Bonett Normed; TLI = Tucker Lewis index; CFI = comparative fit index; IFI = Bollen; MFI = McDonald; GFI = Lisrel GFI; AGFI = Lisrel AGFI; RMR = Root Mean Squared Residual; SRMSR = Standardized Root Mean Squared Residual; RMSEA = Root Mean Square Error of Approximation.

Table F2

Model Evaluation Statistics for the African American Sample without Disabilities Imposed on the African American Sample with Learning Disabilities for the Two, Three, and Four-Factor Model

<u>Goodness of Fit Statistics</u>	<u>Two-Factor Model</u>	<u>Three-Factor Model</u>	<u>Four-Factor Model</u>
χ^2	269.28	214.20	123.52
<u>df</u>	43	43	43
BBN	.88	.85	.76
TLI	.89	.87	.86
CFI	.90	.88	.86
IFI	.90	.88	.87
MFI	.89	.88	.84
GFI	.95	.95	.90
AGFI	.93	.93	.87
RMR	1.48	1.58	1.92
SRMSR	.16	.17	.21
RMSEA	.07	.07	.08
90% Confidence Interval of RMSEA	.07, .08	.06, .08	.06, .10

Note. BBN = Bentler-Bonett Normed; TLI = Tucker Lewis index; CFI = comparative fit index; IFI = Bollen; MFI = McDonald; —GFI = Lisrel GFI; AGFI = Lisrel AGFI; RMR = Root Mean Squared Residual; SRMSR = Standardized Root Mean Squared Residual; RMSEA = Root Mean Square Error of Approximation.

Table F3

Model Evaluation Statistics for the Caucasian Sample with Disabilities Imposed on the African American Sample with Learning Disabilities for the Two, Three, and Four-Factor Model

<u>Goodness of Fit Statistics</u>	<u>Two-Factor Model</u>	<u>Three-Factor Model</u>	<u>Four-Factor Model</u>
χ^2	566.81	414.29	174.88
<u>df</u>	43	52	62
BBN	.75	.71	.66
TLI	.75	.72	.73
CFI	.77	.73	.75
IFI	.77	.73	.75
MFI	.77	.75	.72
GFI	.91	.92	.83
AGFI	.89	.90	.85
RMR	5.38	5.42	5.63
SRMSR	.59	.59	.62
RMSEA	.11	.11	.10
90% Confidence Interval of RMSEA	.10, .12	.10, .11	.09, .12

Note. BBN = Bentler-Bonett Normed; TLI = Tucker Lewis index; CFI = comparative fit index; IFI = Bollen; MFI = McDonald; GFI = Lisrel GFI; AGFI = Lisrel AGFI; RMR = Root Mean Squared Residual; SRMSR = Standardized Root Mean Squared Residual; RMSEA = Root Mean Square Error of Approximation.

Table F4

Model Evaluation Statistics for the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Three-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AAN</u>	<u>CD</u>	<u>AALD</u>
χ^2	114.84	178.25	200.20
<u>df</u>	52	52	52
BBN	.92	.88	.86
TLI	.85	.91	.88
CFI	.96	.91	.89
IFI	.96	.91	.89
MFI	.91	.74	.89
GFI	.94	.84	.95
AGFI	.93	.79	.94
RMR	.71	4.04	1.57
SRMSR	.08	.29	.18
RMSEA	.06	.11	.07
90% Confidence Interval of RMSEA	.05, .07	.09, .13	.06, .08

Note. AAN = WISC-III African American Sample Without Disabilities; CD = WISC-III Caucasian Validity Sample With Disabilities; AALD = African American Sample With Learning Disabilities; BBN = Bentler-Bonett Normed; TLI = Tucker Lewis index; CFI = comparative fit index; IFI = Bollen; MFI = McDonald; GFI = Lisrel GFI; AGFI = Lisrel AGFI; RMR = Root Mean Squared Residual; SRMSR = Standardized Root Mean Squared Residual; RMSEA = Root Mean Square Error of Approximation.

Table F5

Model Evaluation Statistics for the Caucasian Sample without Disabilities Imposed on Other Samples in a Fully Restricted Factor Analysis for the Four-Factor Model

<u>Goodness of Fit Statistics</u>	<u>AAN</u>	<u>CD</u>	<u>AALD</u>
χ^2	113.37	181.62	106.11
<u>df</u>	62	62	62
BBN	.93	.89	.79
TLI	.97	.92	.90
CFI	.97	.93	.90
IFI	.97	.93	.90
MFI	.93	.75	.88
GFI	.95	.85	.91
AGFI	.93	.81	.89
RMR	.65	4.03	.81
SRMSR	.07	.30	.20
RMSEA	.05	.10	.07
90% Confidence Interval of RMSEA	.04, .06	.08, .11	.04, .09

Note. AAN = WISC-III African American Sample Without Disabilities; CD = WISC-III Caucasian Validity Sample With Disabilities; AALD = African American Sample With Learning Disabilities; BBN = Bentler-Bonett Normed; TLI = Tucker Lewis index; CFI = comparative fit index; IFI = Bollen; MFI = McDonald; GFI = Lisrel GFI; AGFI = Lisrel AGFI; RMR = Root Mean Squared Residual; SRMSR = Standardized Root Mean Squared Residual; RMSEA = Root Mean Square Error of Approximation.

APPENDIX G

STANDARDIZED RESIDUALS REPRODUCED FOR THE TWO, THREE, AND
FOUR FACTOR MODELS FOR THE AFRICAN AMERICAN SAMPLE WITH
LEARNING DISABILITIES

Standardized Residuals Reported for the Two, Three, and Four-Factor Models for the
African American Sample with Learning Disabilities

Table G1

Largest Standardized Residuals Reported From EQS for the Two-Factor Model for the African American Sample with Learning Disabilities in Factor Loadings for the Caucasian Sample with Disabilities in a Fully Restricted Factor Analysis

<u>Largest Standardized Residuals</u>				
V 10, V 1	V 10, V 8	V 8, V 8	V 10, V 10	V 1, V 1
-0.87	-0.82	-0.81	-0.81	-0.81
V 8, V 1	V 10, V 6	V 8, V 4	V 4, V 1	V 6, V 4
-0.78	-0.77	-0.76	-0.73	-0.71
V 10, V 2	V 8, V 6	V 10, V 4	V 9, V 1	V 6, V 1
-0.69	-0.69	-0.69	-0.67	-0.67
V 9, V 8	V 8, V 5	V 6, V 5	V 10, V 5	V 4, V 4
-0.67	-0.66	-0.66	-0.65	-0.64

Note. V 1 = Arithmetic; V 2 = Block Design; V 3 = Coding; V 4 = Comprehension;
V 5 = Information; V 6 = Object Assembly; V 7 = Picture Arrangement; V 8 =
Picture Completion; V 9 = Similarities; V 10 = Vocabulary.

Table G2

Distribution of Standardized Residuals for the African American Sample with Learning Disabilities in the Structures and Loadings of the Caucasian Sample without Disabilities in a Fully Restricted Factor Analysis in the Three-Factor Model

	<u>Range</u>	<u>Frequency</u>	<u>Percent</u>
1.	-0.5 - ---	0	0.00%
2.	-0.4 - -0.5	0	0.00%
3.	-0.3 - -0.4	2	3.03%
4.	-0.2 - -0.3	17	25.76%
5.	-0.1 - -0.2	28	42.42%
6.	0.0 - -0.1	11	16.67%
7.	0.1 - 0.0	8	12.12%
8.	0.2 - 0.1	0	0.00%
9.	0.3 - 0.2	0	0.00%
A.	0.4 - 0.3	0	0.00%
B.	0.5 - 0.4	0	0.00%
C.	++ - 0.5	0	0.00%
Total		66	100.00%

Note. Reproduced from EQS data output

Table G3

Largest Standardized Residuals Reported From EQS for the Three-Factor Model for the African American Sample with Learning Disabilities in Factor Loadings for the Caucasian Sample with Disabilities in a Fully Restricted Factor Analysis

<u>Largest Standardized Residuals</u>				
V 1, V 1	V 11, V 1	V 9, V 1	V 11, V 11	V 4, V 1
-1.07	-1.01	-0.94	-0.86	-0.84
V 7, V 1	V 10, V 1	V 11, V 9	V 9, V 9	V 11, V 4
-0.82	-0.77	-0.76	-0.74	-0.73
V 11, V 7	V 11, V 5	V 11, V 10	V 5, V 1	V 6, V 1
-0.73	-0.72	-0.72	-0.70	-0.70
V 9, V 4	V 4, V 4	V 7, V 4	V 10, V 4	V 6, V 4
-0.68	-0.68	-0.67	-0.66	-0.65

Note. V 1 = Arithmetic; V 2 = Block Design; V 3 = Coding; V 4 = Comprehension; V 5 = Digit Span; V 6 = Information; V 7 = Object Assembly; V 8 = Picture Arrangement; V 9 = Picture Completion; V 10 = Similarities; V 11 = Vocabulary.

Table G4

Largest Standardized Residuals Reported From EQS for the Four-Factor Model for the African American Sample with Learning Disabilities in Factor Loadings for the Caucasian Sample with Disabilities in a Fully Restricted Factor Analysis

<u>Largest Standardized Residuals</u>				
V 1, V 1	V 12, V 1	V 9, V 1	V 2, V 1	V 12, V 12
-1.15	-1.05	-0.95	-0.94	-0.92
V 12, V 2	V 5, V 1	V 12, V 5	V 11, V 9	V 4, V 1
-0.92	-0.89	-0.85	-0.81	-0.80
V 9, V 9	V 7, V 1	V 12, V 9	V 8, V 1	V 9, V 2
-0.76	-0.76	-0.76	-0.76	-0.76
V 12, V 7	V 10, V 1	V 7, V 5	V 10, V 2	V 9, V 8
-0.75	-0.73	-0.73	-0.72	-0.72

Note. V 1 = Arithmetic; V 2 = Block Design; V 3 = Coding; V 4 = Comprehension; V 5 = Digit Span; V 6 = Information; V 7 = Object Assembly; V 8 = Picture Arrangement; V 9 = Picture Completion; V 10 = Similarities; V 11 = Symbol Search; V 12 = Vocabulary.

APPENDIX H
ORIGINAL CORRELATION MATRIX AND STANDARD DEVIATIONS FOR
THE AFRICAN AMERICAN SAMPLE WITH LEARNING DISABILITIES

Original Correlation Matrix and Standard Deviations for the African American
Sample with Learning Disabilities

	<u>Arith</u>	<u>Block</u>	<u>Code</u>	<u>Comp</u>	<u>Info</u>	<u>Object</u>	<u>PArr</u>	<u>PComp</u>	<u>Sim</u>	<u>Voc</u>
Arith	1.00									
Block	.27	1.00								
Code	.20	.21	1.00							
Comp	.41	.17	.12	1.00						
Info	.36	.15	.11	.35	1.00					
Object	.22	.45	.16	.16	.11	1.00				
Parr	.21	.22	.13	.26	.20	.29	1.00			
Pcomp	.22	.35	.14	.23	.21	.40	.30	1.00		
Sim	.39	.09	.03	.48	.45	.19	.24	.26	1.00	
Voc	.40	.17	.16	.57	.46	.22	.27	.29	.55	1.00
<u>SDs</u>	2.47	3.78	3.20	2.92	2.99	3.09	4.26	2.95	2.83	2.63

Note. N = 976

ABSTRACT

GENERALIZABILITY OF THE FACTOR STRUCTURE OF THE WISC-III FROM
STANDARDIZATION SAMPLES TO AFRICAN AMERICAN STUDENTS WITH
LEARNING DISABILITIES

Sharon Anne Shindelman, PhD

Fordham University, New York, 2000

Mentor: James J. Hennessy, PhD

This study examined the factor structure of the WISC-III for a large sample of school-age African American students designated as having learning disabilities (AALD) and compared it to the nondisabled Caucasian (CN) and African American (AAN) samples of the WISC-III standardization sample, and the Caucasian disabled (CD) sample of the WISC-III standardization sample. The study used cross-validation of covariance structure models (CVCSM), a statistical technique employing both partially restricted and fully restricted factor analyses to test factor structure generalizability across different samples. Generalizability of the factor structure across the AALD sample was indicated when the goodness of fit of the fully restricted factor analytic models generated from partially restricted factor analyses of the WISC-III standardization samples did not degrade when applied to it. When goodness of fit degraded inconsistently, data was analyzed to determine whether race or disability status or both accounted for the discrepancy in model fit.

The factor structure of the models generalized remarkably consistently across both the WISC-III samples without disabilities (CN and AAN). Only in the two-factor solution did the factor structure generalize adequately across the CD and AALD samples, however. Greater stress in every model application was noted from samples without disabilities to samples without disabilities to samples with disabilities and from the CD samples to the AALD sample. Additionally, the conclusion reached from examination of fit statistics and model evaluation statistics was that the two-factor model provided the most adequate fit across models.

Limitations in the study were addressed. Particular emphases upon the effect of the small sample size of the AALD sample in the four-factor solution. Additionally, the lack of regional representation from every geographical region of the U.S., and the greater representation of lower socioeconomic status families than was reflected in the WISC-III standardization sample were noted. This study suggests future research should use CVCSM in cross validation investigations involving larger samples of the AALD children for whom all 12 subtests scores in the four-factor solution are available. Additionally, it would be interesting to see if “homogeneous” CD samples reflect different factor structures.

VITA

VITA

SHARON ANNE SHINDELMAN

Date of Birth	December 26, 1941
Place of Birth	Bellingham, Washington
High School	Bellingham High School Bellingham, Washington Graduated June 1959
Bachelor of Science Psychology	Seattle University Seattle, Washington Conferred May 1963
Master of Science Clinical Psychology	Fordham University Bronx, New York Conferred February 1966
Doctor of Philosophy Urban School Psychology	Fordham University New York, New York
Academic Positions	Adjunct Professor School Psychology Program Long Island University Tarrytown, New York 1993-1996
	Adjunct Professor School Psychology Program College of New Rochelle New Rochelle, New York 1993-1995
Present Position	School Psychologist and Behavioral Consultant, Bellingham Public Schools Bellingham, Washington