

ABSTRACT

KROUSE, HAILEY ELIZABETH. The Structure of Intelligence of Deaf and Hard of Hearing Children: A Factor Analysis of the WISC-IV. (Under the direction of Jeffery P. Braden).

The present study investigated the factor structure of intelligence of D/HOH children with respect to the CHC theory using the WISC-IV. Measurement invariance of a three-factor WISC-IV model (VCI, PRI and PSI) was tested between the D/HOH sample ($N = 134$) and norm group. Results supported configural invariance between the two groups. Metric invariance was not established for the VCI subtests between groups. However, metric invariance was established for the PRI and PSI subtests (separately) between the D/HOH sample and norm group. Additional tests of scalar invariance were conducted for the PRI and PSI subtests (separately); however, the data did not support the scalar invariance for these subtests across groups. Comparisons of mean scores showed that the mean scores for the VCI ($M = 80.05$), PRI ($M = 96.18$) and PSI ($M = 94.16$) for the D/HOH sample were significantly lower than the mean scores for the norm group ($M = 100$), which were thought to represent population values ($p < .05$). Pearson Product Moment correlations were calculated among the WISC-IV subtests. Of the 28 correlations, 25 were significantly greater than zero (i.e., 95% confidence interval did not contain zero). Overall, the CHC theory (as expressed by the WISC-IV) is an appropriate model of intelligence for D/HOH children. The WISC-IV is thought to be a clinically relevant tool to use in the evaluation of D/HOH children. The PRI continues to be the best estimate of g for D/HOH children.

© Copyright 2011 by Hailey Elizabeth Krouse

All Rights Reserved

The Structure of Intelligence of Deaf and Hard of Hearing Children:
A Factor Analysis of the WISC-IV

by
Hailey Elizabeth Krouse

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Psychology

Raleigh, North Carolina

2012

APPROVED BY:

Jeffery P. Braden, Ph.D.
Committee Chair

Christopher B. Mayhorn, Ph.D.

Adam W. Meade, Ph.D.

John L. Nietfeld, Ph.D.

DEDICATION

My dissertation is dedicated to my grandmother, Verna Maroney Reintjes, who has been there for me every step (and page) of the way. This accomplishment would not have been possible without her support, love and guidance. To quote an old Irish proverb, “when I count my blessings, I count you twice.” I love you.

BIOGRAPHY

Born and raised in Morehead City, North Carolina, Hailey Elizabeth Krouse is the daughter of loving parents, Charles and Anne Krouse, and older sister to two wonderful siblings, Laura and Stephen. Hailey graduated in the top 10 of her class from West Carteret High School in 2000. She then followed in her mother's footsteps and attended the University of North Carolina at Chapel Hill where she majored in Psychology and minored in Spanish. After receiving her Bachelor of Arts degree in May 2004, Hailey was accepted into North Carolina State University's Ph.D. school psychology program where she had the pleasure of working under the direction of Dr. Jeffery Braden (now Dean Braden). She received her Master of Science degree in August 2008 and completed her internship with the Johnston County School District during the 2009-2010 school year. She will receive her Doctor of Philosophy degree in May 2012. (Twelve has always been her lucky number!) Hailey currently resides in Raleigh, North Carolina with her amazing fiancé, Jason, and works for the Wake County Public School System.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Jeff Braden, whose patience and support made this accomplishment possible. His time and knowledge have been invaluable during this learning process. He has challenged me to think critically and has helped me grow as both a researcher and practitioner. I know the skills and lessons he has taught me will continue to provide guidance in my professional and personal life for many years to come. I would also like to extend my gratitude to Adam Meade for all of his statistical assistance. He went over and above the “call of duty” and I appreciate the extra time and effort he spent helping me complete this project. I would also like to thank John Neitfeld Chris Mayhorn, for their time, insightful feedback and collaboration on this project.

I am indebted to the psychologists from across the country who have so generously dedicated their time and energy collecting and entering data for this project. Truly, without their help this study would not have been possible. I want to say a special “thank you” to Fantasy Lozada for helping me better understand the world of factor analysis. There is strength in numbers and AMOS was a much friendlier place with you by my side. I also want to thank all my grad school girls (you know who you are) for their friendship and support over the past eight years. I am grateful to have you all in my life and I am looking forward to many more fun times in the future!

I am forever grateful to my parents, who have always believed I could succeed in anything I tried. They have unconditionally loved, guided and supported me for the past 30 years. I am so proud to be their daughter. I would also like to thank Laura and Stephen for putting up with such a nerdy big sister. I love you both so much!

Finally, I would like to thank my fiancé, Jason Perez. Our life is filled with so much love and laughter and I feel lucky every day that I wake up next to you. I can't image a better partner. He is a formatting genius and I don't know what I would have done without his calming words and peaceful presence. He inspires me to "live my dreams" and is a constant reminder that love is all around.

TABLE OF CONTENTS

List of Tables.....	x
List of Figures.....	xi
Chapter 1.....	1
Introduction.....	1
Defining Intelligence.....	2
Theories of Intelligence.....	3
Spearman’s g-Factor Theory.....	3
Multiple Factor Theories.....	4
Dichotomous Theories.....	5
Cattell-Horn Gf-Gc Theory.....	6
Carroll’s Three Strata Theory.....	7
Cattell-Horn Gf-Gc Theory vs. Carroll’s Three Strata Theory....	8
CHC Theory	9
Intelligence Testing and Theory.....	13
Wechsler Scales of Intelligence and Theory	13
CHC Theory and Cognitive Assessment.....	16
Overview of Factor Analysis.....	17
CHC Theory and the WISC-IV.....	21
Unique Aspects of the D/HOH Population.....	26
Cognitive Abilities of D/HOH Individuals.....	27
Cognitive Differences Between D/HOH and Hearing Individuals.	28

Research Investigating the Intelligence of D/HOH Individuals.....	31
Research Investigating the Structure of Intelligence of D/HOH Indi.....	34
Summary.....	37
Measurement Equivalence/Invariance.....	38
ME/I Model Fit.....	41
ME/I and the WISC-IV.....	41
Purpose of the Present Study.....	42
Hypotheses.....	43
Hypothesis One: Configural Invariance.....	43
Hypothesis Two: Metric Invariance.....	44
Hypothesis Three: Differences in Mean Scores.....	44
Hypothesis Four: Subtest Intercorrelations.....	44
Chapter 2.....	47
Method.....	47
Study Changes.....	47
The WM Subtests of the WISC-IV and D/HOH Examinees.....	47
Participants and Examinees.....	50
Recruitment.....	50
Procedure.....	51
Demographic Data.....	52
WISC-IV Data.....	54
Predictions and Analyses.....	55

Sample Size.....	55
Merging Samples.....	56
Converting Data.....	57
Factor Model.....	58
Factor Details.....	60
Chapter 3.....	63
Analyses and Results.....	63
Participants and Examinees.....	63
Descriptive Analyses.....	70
Preliminary Analysis: Restriction of Range.....	71
Hypotheses.....	72
Chapter 4.....	83
Discussion.....	83
Hypothesis One: Configural Invariance.....	84
Hypothesis Two: Metric Invariance.....	87
Additional Analyses.....	93
Hypothesis Three: Differences in Mean Scores.....	97
Hypothesis Four: Subtest Intercorrelations.....	106
Summary.....	108
Implications for Practice.....	110
Limitations.....	113
Directions for Future Research.....	116

Conclusions.....	118
References.....	120
Footnotes.....	140
Appendices.....	141
Appendix A.....	142
Appendix B.....	144
Appendix C.....	147
Appendix D.....	148
Appendix E.....	149
Appendix F.....	151

LIST OF TABLES

Table 1	Description of the CHC Broad Abilities.....	12
Table 2	The Subtests the Comprise the WISC-IV Indexes.....	22
Table 3	The WISC-IV Subtests and Their Loadings onto Indexes and CHC Broad Abilities.....	23
Table 4	Exploratory Factor Loadings of the WISC-IV Subtests.....	25
Table 5	General Background and Hearing Loss Data Gathered for Each D/HOH Child.....	53
Table 6	Setting and Test Administration Data Gathered for Each D/HOH Child.....	54
Table 7	WISC-IV Data Gathered for Each D/HOH Child.....	55
Table 8	ANOVA Source Table.....	57
Table 9	AFI Model Fit Cutoff Criteria.....	62
Table 10	Demographic Information.....	64
Table 11	Demographic Information Describing Examinees' Hearing Impairment.....	69
Table 12	The Mean and Standard Deviation of the WISC-IV Subtests and Composite Scales.....	70
Table 13	Scalar Invariance Intercept Values.....	78
Table 14	Results from Multiple Tests of Invariance	79
Table 15	Correlations among WISC-IV subtests (with 95% Confidence Intervals reported.....	82

LIST OF FIGURES

Figure 1	The CHC Theory of Intelligence.....	11
Figure 2	Path Diagram of Study Hypotheses.....	46
Figure 3	Factor Model of the WISC-IV.....	59
Figure 4	Standardized estimates for the norm group.....	73
Figure 5	Standardized estimates for the D/HOH sample.....	74

CHAPTER 1

Introduction

Psychologists have been studying intelligence for more than 150 years. For most of this time, the areas of the science dedicated to generating intelligence theory and conducting intelligence testing have remained mutually exclusive. A large gap has existed between intelligence research and practice. Before the 21st century, few intelligence tests were guided by theoretical principles. However, since the birth of the Cattell-Horn-Carroll (CHC) theory of intelligence in 2000, the gap between theory and practice has narrowed. Many recently published intelligence tests, including the Wechsler Intelligence Scales for Children—Fourth edition (WISC-IV), are based on the CHC theory of intelligence. Although evidence exists supporting this model of intelligence for children and adults, there is no research that collects evidence using current tests built on the CHC theory with deaf and hard-of-hearing (D/HOH) populations₁. Therefore, the purpose of this study is to examine the structure of intelligence of D/HOH children using the WISC-IV and to answer the question: Is the CHC theory, as measured with the WISC-IV, a parsimonious model of intelligence for this unique group?

To justify and provide a context for this goal, I will discuss the current definitions and theories of intelligence leading up to the CHC theory. Next, I will provide a discussion of the link between theory and assessment with a concentration on the CHC theory and the new WISC-IV. After that, the focus switches to D/HOH children. First, characteristics and terms unique to this special population are described followed by an overview of the research on the nature and distribution of the intelligence for this special group. After that, a summary of the previous research on the structure of intelligence of D/HOH children is provided

followed by a discussion of measurement invariance methods. Finally, the purpose of the current study is described followed by the hypotheses, methods for data collection, and proposed analyses.

Defining Intelligence

The word *intelligence* comes from the Latin word *intelligere*, which means to *understand*. Although the English word “intelligence” has a definite meaning, the definition of the construct has been an issue of debate for over 100 years (Wasserman & Tulsky, 2005). During the first half of the 20th century, psychologists created numerous definitions of intelligence and there was little unity in defining this construct. Charles E. Spearman (1927) stated, “In truth, ‘intelligence’ has become a mere vocal sound, a word with so many meanings that it finally has none” (p. 14).

The first text to use the term intelligence was Herbert Spencer’s *The Principles of Psychology* originally published in 1855 and revised in 1885. Spencer defined intelligence as the biological ability for an organism to adapt to its environment. Decades later, Binet and Simon (1911; 1916) and Stern (1912; 1914; cited in Mackintosh, 1998) continued to emphasize the importance of adaptation in defining an individual’s intellectual ability. Over time, however, the definitions of intelligence continued to multiply and expand, involving several other cognitive processes (e.g., capacity to learn, sensation, perception, quickness, imagination, span of attention, memory, judgment, reasoning, discrimination, association, understanding abstract concepts). In 1921, a conference was held by the some of the most prominent psychologists in the field (e.g., Wechsler, Thorndike, Thurstone, Terman) with the intent of reaching a consensus on the definition of intelligence. However, their efforts were

unsuccessful. Over half a century later, Sternberg and Detterman (1986) published a follow-up to the 1921 conference. Twenty-five contemporary psychologists were involved; yet again, a common definition was not endorsed (Wasserman & Tulsky, 2005).

Although psychologists do not agree on a single definition of intelligence, one of the most enduring and popular definitions of intelligence was proposed by Wechsler (1939).

Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment. It is global because it characterizes the individual's behavior as a whole; it is aggregate because it is composed of elements or ability, which, though not entirely independent, are qualitatively differentiable (p. 3).

Theories of Intelligence

Psychologists not only struggled with defining intelligence, but also labored over the development of intelligence theories. Interested in using quantitative methods of investigation, psychologists began proposing various models of the structure of intelligence. As the goal of all theory, these psychologists hoped that their theoretical models would create inclusive definitions and valid measures of this construct of interest (i.e., intelligence).

Spearman's g-Factor Theory

In 1904, Spearman developed the first empirical theory of intelligence, which he published in his paper entitled "General Intelligence, Objectively Determined and Measured" (Wasserman & Tulsky, 2005). In this paper, Spearman argued that there is a single, underlying process of general intelligence ("g"), which influences all intellectual activities (Mackintosh, 1998). Spearman noted that when large numbers of individuals are

assessed with a variety of intelligence tests, those that scored highly on one test tend to make high marks on the other tests as well. This observation also held for individuals who made intermediate and low scores (Wechsler, 1974a). Spearman showed that g is a mathematically derived factor of shared variance (Wasserman & Tulsky). Many psychologists describe Spearman's theory as a single factor theory; however, this is incorrect (Flanagan, McGrew & Ortiz, 2000). Spearman's theory, originally called the *two-factor theory*, dichotomized variance into shared variance, common across measures (g) and unique variance, specific to individual tests (s). Although most psychologists do not agree with Spearman's theory, many feel that his discovery is one of the most important contributions to psychology (Wasserman & Tulsky). Wechsler (1939; Wechsler, 1974a) stated "Professor Spearman's generalized proof of the two-factor theory of human abilities constitutes one of the great discoveries of psychology" (p. 35). "Even the most ardent critics of Spearman's work seem unable to totally dismiss the existence of a general factor" (Wasserman & Tulsky, p. 16).

Multiple Factor Theories

The dissection of Spearman's g -factor model began in the early 20th century with Thurstone's use of multiple-factor analysis (Wasserman & Tulsky, 2005). Factor analysis uses the correlations among variables to estimate the number of latent traits or factors within the correlation matrix (Carroll, 2005). When discussing factor analysis, it is important to make the distinction between factors and cognitive abilities. Factors (i.e., the operational indicators) are mathematically derived constructs accounting for observed correlations; whereas, the cognitive abilities are what one infers causes factors (i.e., the common cognitive skill that underlies performance on different tests). Although factors and cognitive abilities

are sometimes treated as synonyms, they are conceptually distinct and should not be confused.

The then-new statistical process of factor analysis held the promise of uncovering the abilities underlying intelligence, thus revealing its true structure. In 1934, Thurstone conducted a centroid factor analysis of a battery of cognitive tests that revealed 13 factors, 7 of which he believed to be primary mental abilities: spatial visualization, perceptual speed, numerical facility, verbal comprehension, associative memory, word fluency, and reasoning. Initially, Thurstone did not find evidence of Spearman's general intelligence, or *g*, factor. However, later, Thurstone developed higher-order factor-analytic techniques where he acknowledged Spearman's *g*-factor as a higher-order factor (Wasserman & Tulsky, 1998).

Thurstone's use of multiple-factor analytic approaches to understanding intelligence allowed psychologists to develop empirically based theories of intelligence, relying on data rather than educated hypotheses. Most subsequent theories of intelligence used the statistical method of factor analysis in understanding the construct of intelligence.

Dichotomous Theories

Cattell worked with both Spearman and Thorndike investigating the factor structure of intelligence and developed one of the most prominent dichotomous models of intelligence. In 1941, Cattell introduced his theory of intelligence, arguing that Spearman's unitary *g*-factor theory of intelligence underrepresented the complex nature of intellectual abilities. Cattell posited that there were two separate general abilities defined by related factors: *Gf*, or fluid intelligence, and *Gc*, or crystallized intelligence. Fluid intelligence, as defined by Cattell (1963; 1971) and later Horn (1976), is most synonymous with the ability to use

reasoning skills (e.g., inductive and deductive) in adapting to novel situations. Cattell believed this to be the most essential feature (or characteristic) of intelligence. Crystallized intelligence refers to the ability to draw on preexisting knowledge and acquire new knowledge using familiar learning strategies. Crystallized intelligence is most often assessed by tapping culturally specific and educationally relevant knowledge such as factual information, word knowledge, quantitative skills, and language comprehension (Wasserman & Tulskey, 2005). Thus, the original *Gf-Gc* theory was a dichotomization of cognitive ability (Flanagan et al., 2000) that posited two related but distinct cognitive abilities.

Cattell-Horn *Gf-Gc* Theory

Cattell and his student, Horn, continued to modify and improve the *Gf-Gc* theory. In the mid-1960s, Horn (1965) expanded the *Gf-Gc* model to include four additional abilities: visual processing (*Gv*), short-term memory (*Gsm*), long-term retrieval (*Glr*) and processing speed (*Gs*). A few years later, Horn (1968) added another factor, auditory processing (*Ga*) and redefined *Gv*, *Gs* and *Glr*. In the 1990s, Horn (1991) added a sixth factor, correct decision speed (*CDS*). A few years later, the final two factors, quantitative ability (*Gq*) and broad reading/writing ability (*Grw*), were added based on the work of Horn (1991) and Woodcock (1994), respectively. This multiple factor model became known as the Cattell-Horn *Gf-Gc* theory (Alfonso, Flanagan, & Radwan, 2005). Although this new theory included the addition of eight new factors (for a total of 10), it was still referred to as the *Gf-Gc* theory. The name causes some confusion because it implies a two-factor structure, whereas it has actually come to be defined as a multiple factor theory (Flanagan, et al., 2000).

Carroll's Three Strata Theory

Drawing on a number of preexisting theories (e.g., Spearman, 1927; Thurstone, 1938; Horn & Cattell, 1966; Horn, 1991), Carroll (1993) devised a theory of intelligence that differentiated cognitive factors into a hierarchical three stratum model. Carroll's three strata, I, II and III, differ in breadth and generality (Carroll, 2005).

Carroll's stratum I factors explain variance that are shared by a few tasks, meaning they rely on a narrow set cognitive abilities (i.e., the abilities are not common across many cognitive tasks). Carroll identified 65 distinct stratum I factors. The stratum II factors explain variance that are shared by a number of first stratum factors, meaning they rely on a more general set of cognitive abilities. Carroll recognizes the following eight stratum II factors: fluid intelligence (*Gf*), crystallized intelligence (*Gc*), general memory and learning (*Gy*), broad visual perception (*Gv*), broad auditory perception (*Gu*), broad retrieval ability (*Gr*), broad cognitive speediness (*Gs*) and decision/reaction time speed (*Gt*). At the highest level or stratum III is *g*, or general intelligence, which encompasses all second stratum factors. Carroll's concept of *g* is consistent with Spearman's (1927) definition of *g* (Alfonso et al., 2005; Carroll, 2005).

In Carroll's three stratum theory, the higher strata subsume the lower strata. For example, general sequential reasoning is a narrow, stratum I factor that falls under the broad, stratum II factor of fluid intelligence (which also includes induction, quantitative reasoning and speed of reasoning). In turn, fluid intelligence contributes to the only factor in stratum III, the *g* factor (Alfonso et al., 2005; Carroll, 2005).

Cattell-Horn *Gf-Gc* Theory vs. Carroll's Three Strata Theory

The Cattell-Horn *Gf-Gc* Theory and Carroll's Three Strata Theory are so similar that many have proposed they represent a common, unified theory (called CHC theory after Cattell, Horn, and Carroll), as will be described in the next section. However, there are four important points of difference that should be made between these two theories. The most notable difference is that Carroll's Three Strata theory includes a general ability factor, *g*, at the highest level, stratum III. The Cattell-Horn *Gf-Gc* theory does not include a representation of *g* in the model. A second distinction is that the Cattell-Horn *Gf-Gc* theory includes quantitative knowledge (*Gq*) as a broad ability, whereas Carroll's Three Strata theory includes quantitative reasoning as a narrow, stratum I ability under the broad, stratum II ability of fluid intelligence (*Gf*). The third difference is that the Cattell-Horn *Gf-Gc* theory includes a broad reading/writing factor (*Grw*) in stratum II, whereas Carroll's Three Strata theory includes reading and writing ability as two narrow, stratum I abilities subsumed under the stratum II ability of crystallized intelligence (*Gc*). The fourth and final difference is that Carroll's Three Strata theory includes a broad, stratum II general memory and learning factor (*Gy*) that includes short-term memory, associative memory, meaningful meaning and free-recall memory, whereas the Cattell-Horn *Gf-Gc* theory has separate factors for short-term memory (*Gsm*) and long-term retrieval (*Glr*), the latter of which includes associative memory, meaningful memory and free-recall memory as stratum I abilities (Alfonso, et al., 2005; Flanagan, et al., 2000).

CHC Theory

In an attempt to resolve the differences between the Cattell-Horn *Gf-Gc* theory and Carroll's Three Strata Theory, McGrew (2000; in Flanagan, et al., 2000) proposed an "integrated Cattell-Horn-Carroll *Gf-Gc* model" (p. 28). The integrated model "represents *both* the Cattell-Horn and Carroll models, in their respective splendor" (McGrew, 2005, p. 149). This integrated theory quickly became known as the Cattell-Horn-Carroll (CHC) theory of cognitive abilities as it began seeping into the psychology literature in the early 21st century (Alfonso, et al., 2005).

The CHC theory (see Figure 1, adapted from Flanagan, et al., 2000) is depicted in a three stratum model with *g* or general intelligence as stratum III, 10 broad or stratum II abilities (i.e., Fluid intelligence [*Gf*], Crystallized intelligence [*Gc*], Quantitative knowledge [*Gq*], Reading and writing [*Grw*], Short-term memory [*Gsm*], Visual processing [*Gv*], Auditory processing [*Ga*], Long-term storage and retrieval [*Glr*], Processing speed [*Gs*], and Decision speed/reaction time [*Gt*]) and 75 narrow or stratum I abilities (e.g., language development, listening ability, simple reaction time, memory span). (For a description of the broad abilities please see Table 1.) Interestingly, the CHC theory omits stratum III, general intelligence, in the model. However, this omission is for practical rather than theoretical purposes. CHC theory proponents recognize the existence of a general human ability, or *g*; however, they note it has little practical utility in guiding test development and, therefore, do not include it in the model (Alfonso, et al., 2005; Flanagan, et al., 2000). The CHC theory "omits a *g* or general ability factor, primarily because the utility of the theory [as it is employed in assessment-related disciplines] is in clarifying individual cognitive and

academic strengths and weaknesses, which are understood best through the operationalization of broad [stratum II] and narrow [stratum I] abilities (Alfonso, et al., 2005, p. 188). However, this is a controversial point, as many in the field (e.g., Braden & Shaw, 2009; Jensen, 1997) argue strongly that the practical value from most cognitive tests is mostly (if not entirely) due to the ability of the tests to measure *g*. Although *g* is not typically depicted in visual models of the CHC theory, it is included in Figure 1. In sum, the CHC theory is the culmination of years of research investigating the structure of human intelligence. “The CHC taxonomy is the obvious cognitive cornerstone of a model of human aptitude” (McGrew, 2005).

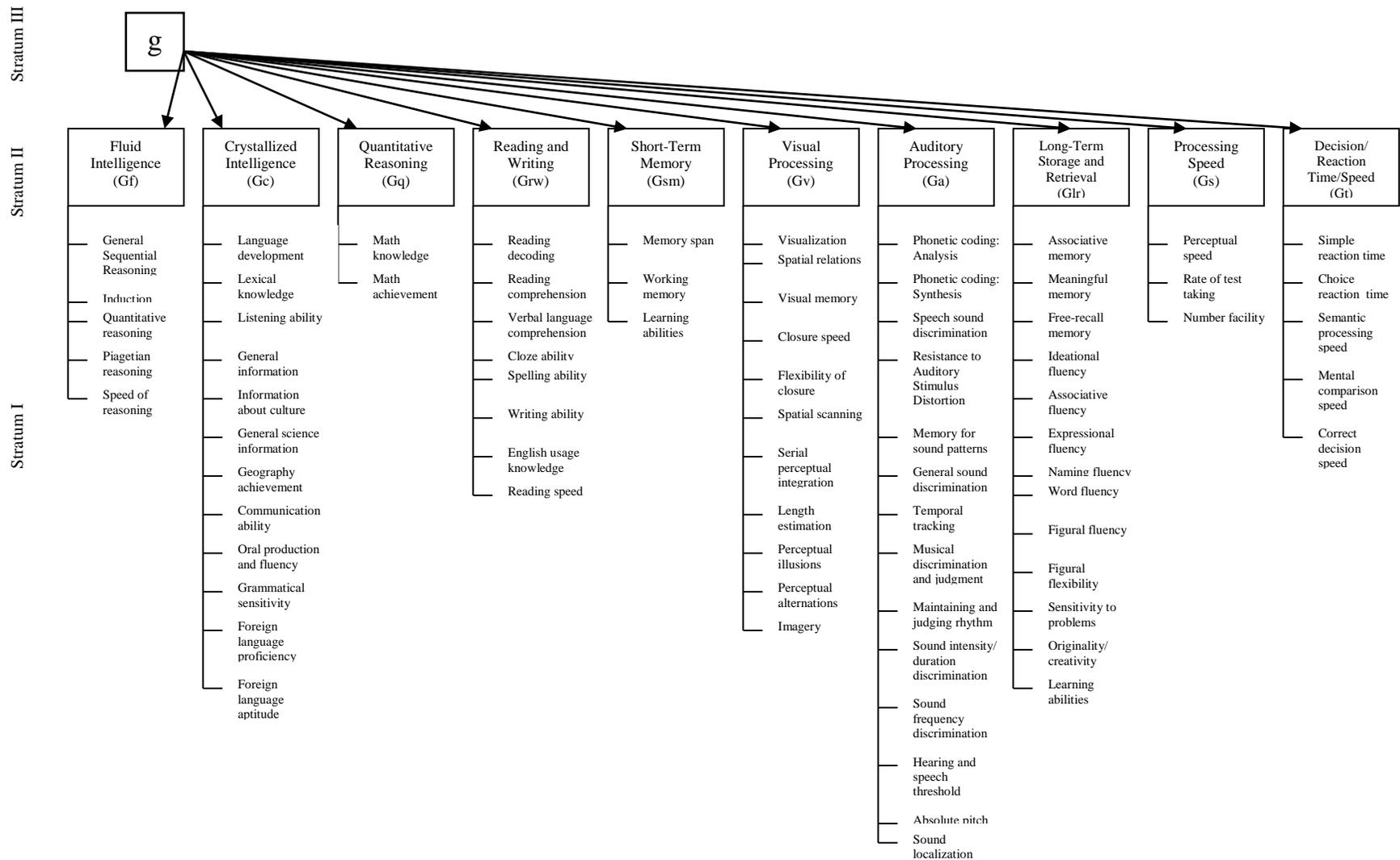


Figure 1. The CHC Theory of Intelligence

Table 1

Description of the CHC Broad Abilities

Broad Ability	Description
Fluid Intelligence (Gf)	Mental operations (e.g., inductive and deductive reasoning) an individual uses to perform novel tasks
Crystallized Intelligence (Gc)	Culturally specific knowledge
Quantitative Knowledge (Gq)	Quantitative declarative and procedural knowledge
Reading and Writing (Grw)	Knowledge of basic reading and writing abilities
Short-term Memory (Gsm)	The ability to mentally hold information and retrieve it within a few seconds
Visual Processing (Gv)	The ability to perceive, analyze, synthesize, generate, store, retrieve, manipulate, and transform visual stimuli
Auditory Processing (Ga)	The ability to perceive, analyze, and synthesize auditory patterns and detect changes in sound patterns
Long-term Storage & Retrieval (Glr)	The ability to store and quickly and accurately retrieve information from long-term memory
Processing Speed (Gs)	The ability to perform cognitive tasks fluently and automatically
Decision/Reaction Time/Speed (Gt)	The ability to react quickly to stimuli

Intelligence Testing and Theory

“The only way we can evaluate intelligence quantitatively is by the measurement of various aspects of these [intellectual] abilities” (Wechsler, 1974a, p. 74). Intelligence tests, from the beginning, have always been viewed as an instrument used to measure the larger construct, intelligence. “...As far as measuring intelligence is concerned, these specific tasks are only a means to an end. Their object is not to test a person’s memory, judgment or reasoning ability, but to measure something which it is hoped will emerge from the sum total of the subject’s [sic] performance, namely, his [sic] general intelligence” (Wechsler, 1974a, p. 35).

Intelligence tests and theory have a reciprocal relationship. Ideally, intelligence theories guide test development; however, intelligence test data are often used to guide theory development. For the majority of the 20th century, intelligence tests largely ignored theory in favor of empiricism. Consequently, there has been a large gap between intelligence theory and intelligence tests. With the exception of the intelligence tests published within the last few decades, most intelligence tests were not based on a specific theory of intelligence. This is true of the Wechsler Scales of Intelligence.

Wechsler Scales of Intelligence and Theory

Wechsler developed his first scale of intelligence in 1939. Since then, Wechsler has created new scales of intelligence for different groups (e.g., adults, children, preschoolers) each with multiple revisions. Over time, the Wechsler Scales of Intelligence have become the most frequently used and researched intelligence tests in the United States. The scales’ clinical utility and psychometric soundness have been well documented. Overall, the

Wechsler Scales of Intelligence have made significant contributions to the fields of clinical and school psychology (Zhu & Weiss, 2005).

With the exception of the WISC-IV (Wechsler, 2003), all other Wechsler Intelligence Scales used the same verbal/performance model of intelligence. “Without a doubt, the Wechsler verbal/nonverbal (performance) model of intelligence is the most widely recognized dichotomous model of cognitive abilities” (Flanagan, et al., 2000, p. 20).

Although Wechsler organized his intelligence tests into two categories, he did not believe in two distinct types of intelligence. Wechsler’s intent was to organize his intelligence tests to reflect the two different methods, or avenues, through which intelligence could be expressed. In other words, Wechsler believed that Spearman’s (1927) concept of general intelligence or *g* could be assessed through multiple modalities, specifically verbal and nonverbal (Flanagan, McGrew & Ortiz, 2000), and that doing so would reduce errors that might occur if only one approach was used.

Although the Wechsler Scales of Intelligence were not based on the dichotomous Cattell-Horn *Gf-Gc* theory of intelligence, many researchers interpreted the Wechsler Scales using this model. “...The Verbal and Performance scales and the IQs they yield were almost universally interpreted as corresponding closely to the *Gc* (Verbal) and *Gf* (Performance) constructs” (Kaufman, 1994, p. 168). In this respect, the difference between an individual’s Verbal IQ (VIQ) and Performance IQ (PIQ) was understood as a discrepancy between fluid and crystallized intelligence rather than the person’s ability to respond to questions verbally or nonverbally (Kaufman, 1994). Unfortunately, the match between Verbal-Performance and Fluid (*Gf*)-Crystallized (*Gc*) dichotomies is imperfect (e.g., Similarities, a verbal subtest

intended to measure G_c , also measures G_f due to its reliance on abstract thought) (Kaufman, 1994).

In the past, efforts to link the Wechsler Intelligence Scale for Children (WISC) to theories of intelligence have been driven by the desire to explain or account for the data. Psychologists (e.g., Bannatyne, Kaufman) developed theories of intelligence based on WISC data and then used the WISC data to confirm the theories. This methodological process succumbs to the tautological fallacy (i.e., using data to drive the model to confirm the model), and fails to provide empirical support for the theory or the test (J.P. Braden, personal communication, December 15, 2008).

As the dichotomous Cattell-Horn G_f - G_c theory of intelligence evolved into a multi-factor model of intelligence, researchers began to evaluate the structure of intelligence tests against this model. “Although there was substantial evidence of at least eight or nine broad cognitive G_f - G_c abilities by the late 1980s, the tests of the time did not reflect this diversity in measurement” (Alfonso, et al., 2005, p. 188).

Driven by theoretical developments independent of the Wechsler Scales, researchers began to investigate the broad cognitive abilities tapped by Wechsler subtests. As a result, researchers began interpreting the WISC-III (Wechsler, 1991) against the multi-factor Cattell-Horn G_f - G_c theory. The analyses revealed that most WISC-III subtests only tapped two broad cognitive abilities, G_c and G_v (consistent with the Verbal-Performance dichotomy). Specifically, with the exception of the Digit Span and the Arithmetic subtests loading onto G_{sm} , the verbal subtests loaded onto G_c , and with the exception of Symbol Search and Coding subtests loading onto G_s , the nonverbal subtests loaded onto G_v . From

this evidence it became clear that Wechsler Scales of Intelligence were not reflecting current theoretical views in measuring the array of cognitive abilities thought to comprise general intelligence, or *g*. "...It is clear that most test authors did not use the theory of fluid and crystallized intelligence and its corresponding research base to guide the development of their intelligence tests. As such a substantial theory-practice gap existed; that is, theories of the structure of cognitive abilities were far in advance of the instruments used to operationalize them" (Alfonso, et al., 2005, p. 192).

CHC Theory and Cognitive Assessment

As CHC theory was accepted as the consensus model for understanding intellectual abilities, test developers began to adopt CHC theory as the foundation on which to base their tests. Overall, the CHC theory has had an impressive impact in the development of intelligence tests. "Nearly every intelligence test developer acknowledges the importance of CHC theory in defining and interpreting cognitive ability constructs, and most have used this theory to guide directly the development of their intelligence tests" (Alfonso, et al., 2005, p. 188). The theoretical soundness and applicability of the CHC theory has allowed it to narrow the gap between research and practice.

The CHC theory has been supported by a number of factor-analytic studies. Research investigating the factor structure of the Woodcock-Johnson—Revised (WJ-R; Bickley, Keith & Wolfe, 1995; Carroll, 2003), Woodcock-Johnson—Third Edition (WJ-III; McGrew & Woodcock, 2001; Taub & McGrew, 2004), Stanford-Binet Intelligence Scales—Fifth Edition (SB5; Roid, 2003) and the Armed Services Vocational Aptitude Battery (ASVAB; Roberts et al., 2000) have found support for a multi-factor structure of intelligence. In addition to

providing evidence for the CHC structure of intelligence, this research also demonstrates that the tests are, in fact, measuring a variety of cognitive abilities consistent with this model (i.e., evidence of test validity based on internal structure).

Overview of Factor Analysis

“Factor analysis is inexorably linked with the development of intelligence theory and intelligence tests. Early intelligence theories and factor-analytic methods were developed in tandem, and the connection continues to this day” (Keith, 2005, p. 581). There are two main types of factor analysis: exploratory factor analysis (EFA), and confirmatory factor analysis (CFA). CFA is typically described as a more theory-driven approach compared to EFA (Keith, 2005). “EFA can be a valuable tool for *developing* theory, whereas CFA may be better suited for *testing* existing theory [italics in original]” (Keith, 2005, p. 582).

With EFA, the researcher holds no specific expectations as to the nature and number of factors that will be extracted from the data. If using EFA, the researcher must make a series of decisions regarding the method of factor extraction and rotation to use and the number of factors to retain. In contrast, CFA requires the researcher to specify, in advance, the expected factor structure including (a) the number of factors, (b) which variables will reflect given factors, and (c) whether the factors are correlated (Thompson, 2004).

In both EFA and CFA, the researcher is interested in investigating the pattern and degree of the factor loadings. Factor loadings are the correlation coefficients/regression weights relating the measured variables (sometimes referred to as indicators) to their corresponding latent variables (i.e., factor). The value of the factor loading represents the expected change in the observed score on the measured variable (i.e., indicator) per unit

change on the latent factor (Vandenberg & Lance, 2000). The goal, when assigning factor loadings, is to achieve Thurstone's simple structure (i.e., as many loadings as possible to be maximally close to either 1.0 or 0.0). In essence, factor rotation allows a researcher to more evenly distribute variability among factors.

There are two main types of factor rotation, orthogonal and oblique. Orthogonal rotation does not assume the latent factors are correlated, while the oblique rotation allows the factors the freedom to correlate (which is expected among the WISC-IV indexes). When factors are uncorrelated (orthogonal rotation), the factor loadings are the correlation coefficients between the variable and factor, which are equal to the regression weights. However, when the factors are correlated (oblique rotation) the factor loadings are the regression weights (Comrey & Lee, 1992).

Whereas EFA and CFA involve examination of the pattern of factor loadings, CFA is used when a researcher wishes to test the data against *a priori* specified model of factor patterns and/or loadings (often referred to as restricting the factor solution). The results of the CFA provide "fit statistics" which provide feedback as to how well the data "fit" the specified model (Keith, 2005). The actual covariance matrix is used to derive parameters (loadings). These estimated parameters are then used to compute a model-implied covariance matrix which is compared to the actual covariance matrix.

There are a variety of different fit statistics that can be used to interpret the results of a CFA. The chi-square (χ^2) is the most commonly used fit statistic for CFAs. The chi-square allows the researcher to determine the probability that the specified model is an appropriate factor solution for the data. Contrary to most significance testing in psychology, with a CFA,

null results of the chi-square is the desired outcome. A large chi-square value and a small probability (e.g., $p < .05$) suggest that the implied variance-covariance matrix and the actual variance-covariance matrix of the data are significantly different; thus, the specified model is a poor fit of the data. When the chi-square is significant, the proposed factor solution is not an accurate representation of the data. However, a small chi-square value and a large probability (e.g., $p > .05$) suggests that the implied variance-covariance matrix and the actual variance-covariance matrix of the data are not significantly different; thus, the specified model is an acceptable fit for the data (Keith, 2005).

Although the chi-square is the most common fit statistic used to analyze CFAs, it has two notable limitations. First, the chi-square is directly affected by sample size. Large sample sizes can result in significant chi-square statistics (i.e., suggesting the model is a poor fit for the data) even when the implied and actual variance-covariance matrixes are only slightly incongruent. However, small samples almost always result in non-significant chi-square statistics (i.e., suggesting the model is a good fit for the data) even when the implied and actual variance-covariance matrixes are noticeably different. Second, assuming an appropriate sample size, the chi-square statistic tests the exact fit of the data to the model. One small difference between the implied and actual models and the chi-square statistic will be significant, implying a poor fit (Keith, 2005). At best, models are designed to approximate reality and should not be discarded for trivial differences. Therefore, the chi-square statistics may imply the model is inappropriate, when, in fact, it is an appropriate approximation for the data. Because the chi-square has notable limitations, other fit statistics (e.g., Tucker-Lewis Index [TLI], comparative fit index [CFI], adjusted goodness-of-fit index [AGFI], root

mean square error of approximation [RMSEA], standardized root mean square residual [SRMR]) are often used in addition to the chi-square to test the adequacy of the implied factor structure in a CFA (for more information on additional fit indices see Hu & Bentler, 1998).

Although CFA is a common technique used to test the factor structure of intelligence theory and tests, there are limitations to CFA as a statistical tool. CFA is used to test the accuracy of a single model. Results of the fit statistics yield information either in support or rejection of the model. The simple fact that one model does or does not fit the data provides only limited information about the theory being tested. One model may fit the data well, but other models may fit the data better; or, the model may be a poor fit of the data, but is the best option compared to the alternatives. Furthermore, theories may provide support for multiple factor models and a method is needed for finding the best model for the data. Using different approaches, CFA and various fit statistics can be used to compare competing models. The chi-square difference ($\Delta\chi^2$) can be used to test the difference in nested models (used when assessing measurement invariance/equivalence, see Vandenberg & Lance, 2000, pp. 45-47). The Akaike information criterion (AIC) is another method used to compare models with the smaller AIC value suggesting the better model. Similarly, the Bayes information criterion (BIC) and the Browne-Cudeck criterion (BCC) can also be used to find the best fit model among many (Keith, 2005).

In sum, factor analysis is a common statistical procedure used in the development and validation of many intelligence tests and theories. Whereas EFA does not require the researcher to have any a priori expectations as to the structure of the data, CFA is used to test

the “fit” of a specified factor model. Various statistics (e.g., chi-square, TLI, CFI, RMSEA, SRMR) provide information about the “fit” (i.e., good, adequate, poor) of the hypothesized model. Although the fit statistics, as well as CFA, as a statistical tool, have several limitations, the results can provide valuable information about the validity of a test or theory.

CHC Theory and the WISC-IV

The WISC-IV (Wechsler, 2003) included significant modifications to the content and structure of the scale compared to previous editions. The developers of the WISC-IV abandoned the Verbal-Performance structure of earlier editions and elevated the four-factor structure (originally identified in the WISC-III) as the primary method of score interpretation. The four WISC-IV indexes include: (a) Verbal Comprehension Index (VCI), (b) Perceptual Reasoning Index (PRI), (c) Working Memory Index, and (d) Processing Speed Index (PSI). The FSIQ, the best indicator of *g*, was retained as the general composite score due to the fact that it is so widely used in assessments and research (Prifitera, Weiss, Saklofske & Rolfhus, 2005; Saklofske, Rolfhus, Prifitera, Zhu & Weiss, 2005). For an overview of the WISC-IV indexes, please see Appendix A. Table 2 provides a list of the WISC-IV subtests and the Index to which each subtest contributes.

Table 2

The Subtests that Comprise the WISC-IV Indexes

VCI	PRI	PSI	WMI
Similarities	Block Design	Coding	Digit Span
Vocabulary	Picture Concepts*	Symbol Search	Letter Number Seq.*
Comprehension	Matrix Reasoning*	<i>Cancellation*</i>	<i>Arithmetic</i>
<i>Information</i>	<i>Picture Completion</i>		
<i>Word Reasoning*</i>			

Note. New subtests are marked with an *; supplemental subtests are *italicized*.

Although the authors of the WISC-IV do not explicitly state the revisions were rooted in the CHC theory, they acknowledge the research of Cattell, Horn, and Carroll. In looking at the structure of the WISC-IV, the subtests adequately measure five broad CHC abilities (i.e., *Gf, Gc, Gv, Gsm, Gs*). With the exception of the PRI and WMI subtests, which load onto two CHC broad abilities, the subtests of WISC-IV indexes map directly onto the CHC broad abilities (Alfonso, et al., 2005). The WISC-IV, with its theoretical underpinnings and continued practical utility, provide an improved connection between theory (i.e., the CHC theory) and practice. To understand the relationship between the WISC-IV subtests and indexes and between the WISC-IV subtests and CHC broad abilities (as assigned by Alfonso, et al., 2005) please see Table 3.

Table 3

The WISC-IV Subtests and Their Loadings onto Indexes and CHC Broad Abilities

WISC-IV Subtests	WISC-IV Index	CHC Broad Ability
Similarities	VCI	<i>Gc</i>
Vocabulary	VCI	<i>Gc</i>
Comprehension	VCI	<i>Gc</i>
<i>Information</i>	VCI	<i>Gc</i>
<i>Word Reasoning</i>	VCI	<i>Gc</i>
Block Design	PRI	<i>Gv</i>
<i>Picture Completion</i>	PRI	<i>Gv</i>
Matrix Reasoning	PRI	<i>Gf</i>
Picture Concepts	PRI	<i>Gf</i>
<i>Arithmetic</i>	WMI	<i>Gf</i>
Digit Span	WMI	<i>Gsm</i>
Letter-Number Sequencing	WMI	<i>Gsm</i>
Symbol Search	PSI	<i>Gs</i>
Coding	PSI	<i>Gs</i>
<i>Cancellation</i>	PSI	<i>Gs</i>

Note. Supplemental subtests are italicized. VCI = Verbal Comprehension Index. PRI = Perceptual Reasoning Index. WMI = Working Memory Index. PSI = Processing Speed Index. *Gc* = Crystallized Intelligence. *Gv* = Visual Processing. *Gf* = Fluid Intelligence. *Gsm* = Short-term Memory. *Gs* = Processing Speed.

The evidence for the factor structure of the WISC-IV provides evidence for the CHC theory of intelligence. In investigating the factor structure of the WISC-IV, test developers reported EFA and CFA using data from the norm group ($N = 2,200$). Based on a predicted four-factor structure (identified in Table 3), the developers conducted an EFA using the 10 core WISC-IV subtests. Principal axis factoring extraction method was used with an oblique rotation to allow correlations among factors. The results revealed the expected four-factor structure (see Table 4). None of the secondary loadings exceeded .20 (Wechsler, 2003).

Table 4

Exploratory Factor Loadings of WISC-IV Subtests

	Factor			
	Verbal Comprehension (<i>Gc</i>)	Perceptual Reasoning (<i>Gv, Gf</i>)	Working Memory (<i>Gsm</i>)	Processing Speed (<i>Gs</i>)
Similarities	<u>.74</u>	.19	-.03	-.06
Vocabulary	<u>.84</u>	.02	.03	-.02
Comprehension	<u>.78</u>	-.11	.03	.08
Block Design	.01	<u>.66</u>	-.02	.08
Picture Concepts	.13	<u>.45</u>	.03	.03
Matrix Reasoning	.00	<u>.69</u>	.06	.01
Digit Span	.00	.07	<u>.62</u>	-.06
Letter-Number Seq.	.09	-.02	<u>.62</u>	.06
Coding	.02	-.01	-.04	<u>.68</u>
Symbol Search	-.01	.09	.04	<u>.65</u>

Following the exploratory factor analysis, the developers conducted a CFA using data from the norm group. CFA tests the factor structure specified by the researchers *a priori*. For the CFA, the developers tested four different models each with a different factor structure. The results revealed that the four-factor model was the best fit (Wechsler, 2003).

Overall, these data were consistent with predictions based on the model driven by the CHC theory. The findings provide evidence for the reciprocal conclusions that (a) the WISC-IV is consistent with the CHC theory, and (b) the CHC theory is supported by the WISC-IV data. It is important to recognize, however, that most of the research conducted thus far on the CHC theory (with the WISC-IV as well as with other tests of intelligence) has used samples of (normal) hearing children. No research, to date, has employed CFA methods using models based on the CHC theory (e.g., the WISC-IV) to explore the degree to which data from D/HOH samples reflects the CHC factor structure. The question of whether the CHC theory is a parsimonious model of intelligence for D/HOH children is one that is relevant for theoretical and practical reasons, as explained in the following sections.

Unique Aspects of the D/HOH Population

D/HOH children, although commonly grouped together for research purposes, are an extremely heterogeneous group. When describing children with hearing loss, it is important to fully describe the degree and type of hearing loss as well as explain specific aspects of functioning. Researchers should always gather an array of demographic information (e.g., degree of hearing loss, type of hearing impairment, age of onset) about the D/HOH participants to give a comprehensive description of the sample as well as to note specific benefits and/or limitations of the external validity of the study. For a review of the unique

terms and characteristics of this special population as well as prevalence rates, please see Appendix B.

Cognitive Abilities of D/HOH Individuals

More than 75% of the D/HOH children in this country have normally hearing parents and about 70% have normally hearing siblings. In addition, approximately 40% of the D/HOH children in America are born with their hearing impairment and another 14% acquire their hearing loss before the age of two (Gallaudet Research Institute, 2011). Therefore, the majority of D/HOH children in this country have a prelingual hearing loss and are raised in households in which the primary mode of communication is speech. For the majority of prelingually deaf children, exposure to language does not begin until after the diagnosis of deafness and medical and educational interventions have begun, which can be months or even years after the hearing loss occurs. Even after the diagnosis, the degree of language exposure varies greatly depending on the resources available to the family, the communication needs of the child, the skills and training needed to learn alternative modes of communication, and the motivation of the family and child. As a result, most D/HOH children are denied the early and consistent access to language available to normally hearing children. D/HOH children experience fewer, less frequent, and less intense linguistic interchanges, many of which are restricted in content and variety, than their normally hearing peers. Over time, the limited, reduced, and nonstandard exposure to language results in poor language development for these children (Braden, 1994). D/HOH children (especially deaf children) do not acquire and develop spoken language skills in the same manner as normally hearing children (Maller & Braden, 1993).

For D/HOH children, their hearing loss reduces opportunities to observe and participate in oral communication (Braden, 1994). Their inconsistent and infrequent exposure to oral expression reduces their opportunities to learn general and culturally specific knowledge. Their lack of language exposure can adversely affect their intellectual development, especially for cognitive processes dependent on language and audition (i.e., auditory processing [*Ga*] and verbal comprehension and reasoning [*Gc*]; Braden, 2005).

Although some cognitive processes may not be strongly developed for D/HOH individuals, research suggests others (e.g., *Gv*) may develop in a manner and at a level typical of normally hearing children. Some of the observed differences between deaf and hearing individuals are related to the use of the sign language or dependence on visual experience, whereas other differences are due to the lack of auditory encounters (Marschark, 2006).

Cognitive Differences Between D/HOH and Hearing Individuals

For years, researchers have been interested in the cognitive differences between D/HOH and normally hearing individuals. There are practical and theoretical implications for understanding what, if any, qualitative or quantitative differences exist between D/HOH and hearing individuals on various cognitive abilities such as visual-spatial acuity and perception and memory. Pintner and colleagues (e.g., Pintner & Reamer, 1920; Pintner, Eisenson & Stanton, 1941) were among the first to investigate these issues. Through the systematic study of the cognitive abilities of deaf individuals, Pintner et al. concluded that deaf individuals were intellectually inferior to hearing individuals, based on low tests scores across a range of cognitive abilities. However, it should be noted that Pintner's systematic studies included

paper-and-pencil tests requiring deaf participants to respond to verbally loaded questions, confounding language abilities with cognitive abilities. Pintner's flawed methodologies and conclusions led many researchers on a quest to develop nonverbal measures of cognitive abilities appropriate for this special population (Quigley & Paul, 1991).

Pintner's conclusions about the cognitive inferiority of deaf individuals dominated this area of research until the mid-1900s when Mykelbust (1960) challenged Pintner's formulations. Through numerous studies, Mykelbust showed that, when controlling for verbally loaded tasks, deaf and hearing individuals are intellectually equal and score similarly on global measures of intelligence (e.g., WISC Full Scale IQ score). However, Mykelbust also noted specific differences between the cognitive abilities of deaf and hearing individuals (e.g., the two groups produced different score profiles on the WISC subtests). These findings led Mykelbust to conclude that there are quantitative similarities but qualitative differences between the intellectual abilities of deaf and hearing individuals. The observation of these qualitative differences was the basis for Mykelbust's "organismic shift hypothesis," stating that the basic experiences of deaf individuals are altered as a result of their hearing impairment, which, as a consequence, alters subsequently developed behaviors, making deaf and hearing individuals inherently different from one another (Quigley & Paul, 1991).

Mykelbust's theories, however, were quickly overshadowed by belief that deaf and hearing individuals are similar on all important cognitive abilities. Rosenstein (1961), Furth (1966) and Vernon (1967) all concluded that few, if any, differences in cognitive functioning exist between deaf and hearing individuals. It is now generally accepted by many researchers

that what differences are observed between the cognitive abilities of deaf and hearing individuals are quite possibly the result of environmental circumstances or design characteristics rather than inherent differences between the populations (Quigley & Paul, 1991). Quigley and Kretschmer (1982) identified three different types of task influences: “1) the inability of the researcher to properly convey the task demands because of language difference or deficits on the part of the subject, 2) implicit bias within the solution of the task, or 3) general experiential deficits (including verbal language and communication in general) on the part of the subjects” (p. 51).

Yet, despite the possibility that measured cognitive differences between D/HOH and normally hearing individuals may be the consequence of task influences, it is also important to consider environmental influences on deaf individuals. “Beyond the normal heterogeneity seen in the hearing population, differences in the environments and experiences of deaf children and hearing children might lead to different approaches to learning, to knowledge organized in different ways, and to different levels of skill in various domains. Ignoring this possibility not only denies the reality of growing up deaf in a largely hearing world, but jeopardizes academic and future vocational opportunities for deaf children” (Marschark, 2003, p. 464). Most research suggests that deaf and hearing individuals vary in their approaches to cognitive tasks and that these differences are likely influenced by the mode of communication. However, “different” does not mean “deficient.” Exploring the cognitive differences between deaf and hearing individuals will help psychologists and educators make more informed decisions with the goal of optimizing educational and occupational environments for this unique population (Marschark, 2003).

In sum, researchers have moved away from the idea that D/HOH and hearing individuals are inherently deficient in their cognitive functioning. Currently, the majority of researchers believe D/HOH and hearing individuals are cognitively similar, and that their test score differences are likely the result of measurement artifacts. Yet, despite these conclusions, researchers continue to investigate cognitive differences between D/HOH and normally hearing individuals. Numerous studies have been conducted that examine the specific cognitive abilities (e.g., visual-spatial abilities, short-term/working memory) of D/HOH individuals compared to their normally hearing peers.

Research Investigating the Intelligence of D/HOH Individuals

Most of the research investigating the intelligence of D/HOH individuals has focused on differences between language-loaded (i.e., verbal) and language reduced (i.e., nonverbal) tests. Specifically, the majority of the research has explored differences in the means and standard deviations of these two types of intelligence test variations for D/HOH individuals. In considering the CHC theory, verbally loaded tests of intelligence likely tap Gc , and nonverbal tests of intelligence primarily measure a mixture of Gv and Gf abilities. However, past research has not been interpreted in terms of CHC abilities, and current studies have not explored differences among CHC stratum II abilities for D/HOH children.

D/HOH individuals tend to score similarly to hearing individuals on nonverbal IQ tests (Braden, 1990; 1994; Braden, Kostrubala & Reed, 1994; Kostrubala, 1998; Maller & Braden, 1993; Murphey, 1957; as cited in Evans, 1980). On verbally loaded IQ tests, D/HOH children tend to score approximately one standard deviation below the mean of hearing examinees (Braden, 1994, Maller & Braden, 1993). Looking at ipsative scores, D/HOH

children also repeatedly score significantly lower (approximately one standard deviation or 15 points) on verbal tests of intelligence compared to those assessing intelligence nonverbally (Braden, 1994; 2005; Sullivan & Montoya, 1997).

I (Krouse, 2008) conducted a study investigating the psychometric properties of the WISC-IV with D/HOH children. For my sample of D/HOH children ($n = 128$), I found that the mean PRI ($M = 93.21$) and VCI ($M = 80.86$) were significantly lower ($p < .001$) than the mean of the norm group ($M = 100$). However, the mean VCI was not significantly lower than one standard deviation from the mean ($M = 85$). In addition, subtest intercorrelations were assessed among the WISC-IV subtests with Pearson Product Moment correlations. Of the 44 correlations, 29 were significantly greater than zero (i.e., the 95% confidence interval did not contain zero).

Psychologists do not recommend the use of verbal, language loaded, tests (i.e., tests that primarily measure Gc) to measure the general intelligence of D/HOH individuals. Verbal tests confound language skills with intelligence. “There is uniform agreement that systematic deprivation of exposure to verbal, socially specific knowledge impairs performance on verbal scales independent of an individual’s underlying aptitude” (Braden, 1994, p. 76).

According to Messick (1995), construct-irrelevant variance occurs when a test includes factors extraneous to the construct of interest. These irrelevant factors (e.g., oral language skills) either serve to increase or decrease the difficulty of the test for a person or a particular group of people (e.g., D/HOH individuals). For D/HOH people, low scores on a test that measures Gc may be more influenced by their (lack of) opportunity to acquire verbal and social knowledge rather than limited intellectual ability (Braden, 1994; Braden &

Hannah, 1998; Maller, 1996). For D/HOH people, tests of intelligence that measure G_c (but are interpreted as reflecting g) likely tap construct-irrelevant variance (i.e., oral language skills), which will invalidate their meaning as indicators of general intelligence. For this reason, many psychologists and researchers believe that language reduced (or nonverbal) measures of intelligence are more appropriate for estimating general intellectual ability in D/HOH children (Braden, 1994; 2000; Braden & Athanasiou, 2005).

However, the use of nonverbal measures to assess intelligence raises two critical issues. First, it should be noted that very few tests/scales completely eliminate language from directions, content, and responses (McCallum, 2003). Therefore, despite the title of the test or how it “claims” to measure intelligence, only those tests that actually eliminate or drastically reduce the need for language in understanding, processing, and responding to test items are considered truly “nonverbal” (Braden & Athanasiou, 2005).

Second, it is not clear whether nonverbal tests can adequately represent the intended construct of interest (e.g., general intelligence, or g). In other words, does the nonverbal measure of intelligence adequately capture the range of cognitive processes thought to compose “intelligence?” Construct underrepresentation occurs when the test too narrowly samples the construct of interest (Messick, 1995). This is important in the discussion of nonverbal assessment of intelligence because reduction or omission of language-loaded tests (i.e., verbal tests) may reduce the construct of interest (i.e., g). In regards to the CHC theory, two of the stratum II factors, crystallized ability (G_c) and literacy or reading/writing ability (Grw), are strongly related to language. Most nonverbal measures of intelligence exclude these stratum II factors, therefore reducing the representation of abilities thought to be

important in the estimation of *g*. In other words, many nonverbal tests of intelligence exclude important domains of functioning thought to be fundamental to the construct of interest (i.e., *g*) (Braden & Athanasiou, 2005; Ortiz & Dynda, 2005). Using only the performance subtests of a test battery to assess the intelligence of D/HOH individuals may lead to construct underrepresentation. Therefore, the concept of construct underrepresentation is cited as support for using verbal intelligence measures with D/HOH children (Braden & Hannah, 1998; Maller, 1996).

Despite this argument however, there has been abundant research showing that *g* is fairly accurately and easily estimated from even small samples of tests. In other words, *g* is relatively robust, and does not appear to be highly susceptible to underrepresentation. However, CHC stratum II factors may be more vulnerable to underrepresentation than *g*, especially if they were systematically omitted from a test battery due to language loading (Jensen, 1998).

Research Investigating the Structure of Intelligence of D/HOH Individuals

Few studies have investigated the factor structure of intelligence for D/HOH children. Specifically, four studies investigate the factor structure of intelligence for D/HOH populations using the Wechsler Scales as the measure of intelligence. Two studies assessed the factor structure of intelligence using the WISC-R, and two studies used the WISC-III.

Braden (1984) investigated the factor structure of the WISC-R Performance Scales for a sample of deaf children ($n = 1228$) and compared it the factor structure for hearing children (i.e., the WISC-R standardization sample, $n = 2200$). Braden hypothesized that, like the norm group, only one factor would emerge among the five core WISC-R Performance

Scale subtests (i.e., Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding) for his deaf sample. Overall, the results supported his hypothesis. Only one factor emerged for the deaf sample and Braden concluded that the WISC-R Performance Scale measures the same construct (likely comparable to the CHC stratum II *Gv* ability) in the two groups regardless of differences in language exposure. This study supports suggests that the intelligence has the “same character, or structure, in deaf and hearing children” (p. 406).

Expanding upon Braden’s (1984) study, Sullivan and Schulte (1992) compared the factor structure of the full WISC-R (i.e., both Performance and Verbal Scale subtests) with D/HOH ($n = 368$) to the standardization sample reported in the WISC-R manual. Sullivan and Schulte administered the WISC-R to D/HOH and hearing children to assess whether the same three-factor structure (i.e., Verbal Comprehension [likely comparable to the CHC *Gc* ability], Perceptual Organization [likely comparable to the CHC *Gv* ability] and Freedom from Distractibility [likely comparable to the CHC *Gsm* ability]) described in the WISC-R Manual (Wechsler, 1974) emerged for both groups. Sullivan and Schulte conducted separate factor analyses for the deaf ($n = 291$) and hard-of-hearing ($n = 71$) children; however, the factor structure was identical for the two groups and, therefore, the two groups were combined for further analyses.

The results of the factor structure with the D/HOH sample yielded a two-factor structure (i.e., Language Comprehension [likely comparable to the CHC *Gc* ability] and Visual-Spatial Organization [like comparable to the CHC *Gv* ability]). This two-factor structure for the D/HOH sample was not consistent with the three-factor structure for the

WISC-R reported in the manual. For the D/HOH sample, two of the three Freedom from Distractibility subtests (i.e., Arithmetic and Digit Span) loaded onto the Language Comprehension factor and the other subtest (i.e., Coding) loaded onto the Visual-Spatial Organization factor (Sullivan & Schulte, 1992). Overall, the results of this study are inconsistent with Braden's (1984) conclusions that D/HOH and hearing children have the same intelligence structure. This study found a different factor structure for D/HOH and hearing children for the WISC-R Performance Scale subtests. A likely reason for these findings is the subtests measuring Gc for the norm group are not, in fact, measuring Gc for the D/HOH sample. For D/HOH children, Gc is confounded by language abilities and, as a result, is not a valid measure of intelligence.

Slate and Fawcett (1995) investigated the factor structure of the WISC-III Performance Scale for D/HOH children ($n = 47$). Like the previous editions of the WISC, the WISC-III is divided into two scales, the Verbal Scale and the Performance Scale. However, unlike the previous editions, the WISC-III subtests load onto four factors: Verbal Comprehension (likely comparable to the CHC Gc ability), Perceptual Organization (likely comparable the CHC Gv ability), Freedom from Distractibility (likely comparable to the CHC Gsm ability) and Processing Speed (likely comparable to the CHC Gs ability). From the Performance Scale subtests, two factors were extracted for the D/HOH sample, Perceptual Organization (likely comparable to the CHC Gv ability) and a Processing Speed (likely comparable to the CHC Gs ability). The two-factor structure is identical to that of the WISC-III norm group. These results support the conclusion that D/HOH and hearing children have the same structure of intelligence.

Sullivan and Montoya (1997) investigated the factor structure of the WISC-III with 106 D/HOH children, including both the Verbal Scale and Performance Scale subtests. The results revealed two factors: Visual-Spatial Organization (likely comparable to the CHC *G_v* ability), and Language Comprehension (likely comparable to the CHC *G_c* ability). The Freedom from Distractibility (likely comparable to the CHC *G_{sm}* ability) and Processing Speed (likely comparable to the CHC *G_s* ability) factors do not emerge for this special population. Sullivan and Montoya's findings are not consistent with Slate and Fawcett's (1995) findings; however, the two-factor structure that emerged for D/HOH children is consistent with Wechsler's (1939) original belief that two factors underlie intelligence, verbal and nonverbal.

Verbal subtests (i.e., those measuring *G_c*) are confounded by oral language skills (or lack thereof), and are therefore poor or attenuated measures intelligence for this special group. In other words, the factors extracted from Sullivan and Montoya's study represent two unique constructs (i.e., language skills and intelligence) rather than two factors (i.e., *G_c* and *G_v*) from the same construct (i.e., intelligence). Therefore, the dichotomous factor structure of the WISC-III confounds distinct abilities and language learning opportunities in D/HOH populations.

Summary

The factor structure of intelligence reflected by the Wechsler Performance Scales of Intelligence is similar for D/HOH and hearing children. However, research investigating the structure of intelligence for D/HOH children using both performance (or nonverbal *G_v/G_f*) and verbal (i.e., *G_c*) subtests of the Wechsler Scales of Intelligence has found different factor

structures for D/HOH and hearing children. The different factor structures found for D/HOH and hearing children are likely due to the fact that subtests measuring *Gc* are not actually assessing intelligence for D/HOH children; rather, these subtests are confounded by the underdeveloped language skills of this unique population. As a result, subtests measuring *Gc* are not valid for assessing the intelligence of D/HOH children because they are confounded with a distinct underlying construct (i.e., oral language skills).

Overall, however, the previous research investigating the intelligence of D/HOH children have revealed a relatively unitary structure with little differentiation among cognitive abilities. This finding may be representative of the structure of intelligence for this special population or may be an artifact of the structure of the intelligence test (i.e., the previous editions of the Wechsler Scales of Intelligence have not been consistent with current multi-factor theories of intelligence). Investigating the factor structure of the intelligence of D/HOH children using the new WISC-IV will provide valuable information due to the fact that this latest edition of the Wechsler Scales is grounded in CHC theory.

Measurement Equivalence/Invariance

Investigating the similarity of the factor structure and loadings of a measure (e.g., the WISC-IV) for two or more groups (e.g., hearing vs. D/HOH) is referred to as measurement equivalence/invariance (ME/I). “Measurement equivalence/invariance (ME/I) can be thought of as operations yielding measures of the same attribute under different conditions” (Meade & Lautenshlager, 2004a, p. 61).

CFA tests of ME/I involve simultaneously fitting a measurement model to two (or more) data samples (Meade & Bauer, 2007). Vandenberg and Lance (2000) recommended

several ‘best practices’ for testing ME/I using the CFA methodology. First, the authors suggest conducting a test of equivalent factor structures across groups (i.e., test of configural invariance). Configural invariance involves testing the pattern of free and fixed factor loadings among groups (Vandenberg & Lance, 2000). If differences between the factor structures are found, then the measures are said to be inequivalent and no further testing is needed. If, however, the results suggest configural invariance across groups then a more stringent test of ME/I can be conducted (i.e., metric invariance).

The test of metric invariance assesses the equivalence of variable factor loadings across groups (i.e., $\Lambda^x = \Lambda^{x'}$). Metric invariance tests the extent to which the relationships between the and factors are equivalent across the groups. Metric invariance assesses whether the degree to which the items load onto the latent factors is equivalent across groups. Factor loadings correspond to the regression slopes/weights of the observed variable on the latent factor. The value of the factor loading represents the expected change in the observed score on the measured variable per unit change on the latent factor (Vandenberg & Lance, 2000). Higher factor loadings suggest a stronger relationship between the latent factor and the observed variable. In a perfect test (with no error) one unit change in the latent factor would be matched by one unit change in the observed variable. However, all measurement contains some degree of error. There are two main sources of variation in observed scores: (a) latent factors and (b) unique error, which consists of unmeasured latent variables and random error (Bryant & Yarnold, 1995). Establishing metric invariance means that the two groups are interpreting the items in the same way. Lack of metric invariance may imply that some items are more important or salient to the construct for one group then for the other. When factor

loadings differ across groups, tests for partial invariance can be conducted (Vandenberg & Lance). “The logic of testing for partial ME/I is that invariance restrictions may hold for some but not all manifest measures across populations, and relaxing invariance constraints where they do not hold controls for partial measurement inequivalence” (Vandenberg & Lance, 2000, p. 18).

If metric invariance (or partial metric invariance) is established, Vandenberg and Lance recommend testing for scalar invariance. Scalar invariance tests the null hypothesis that intercepts of like items’ regressions on the latent variable(s) are invariant across groups (i.e., $\tau^g = \tau^{g'}$). In other words, scalar invariance tests whether the regression intercepts linking the manifest measures to the underlying construct(s) are invariant across groups. Item intercepts are the values of the observed scores when the latent trait is zero. Scalar invariance implies that the group differences in the means of the observed items are due to differences in the means of the underlying construct. It addresses the question of whether there is consistency between group differences in observed means and latent means (Steenkamp & Baumgartner, 1998). “Even if an item measures the latent variable with equivalent metrics... (i.e., metric invariance), scores on that item can still be systematically upward or downward biased” (Steenkamp & Baumgartner, 1998, p. 80). Comparisons of “biased” means are meaningless, unless the bias is removed from the data. The primary purpose of this validity evidence is to support the claim that the manifest measure (i.e., the construct-valid operationalization of the latent factor) reflects the underlying latent attribute in the same way and to the same degree across groups (Vandenberg & Lance). “Thus, demonstration of measurement equivalence is a logical prerequisite to the valuation of substantive hypotheses

regarding group differences, regardless of whether the comparison is as simple as a between-group mean differences test or as complex as testing whether some theoretical structural model is invariant across groups” (Vandenberg & Lance, 2000, p. 9).

ME/I Model Fit

Overall, ME/I model fit refers to evaluating whether the factor model underlying a specific measure or scale is equivalent across different groups. Good model fit provides support that the factor model is invariant across groups whereas poor model fit suggests significantly different factor models across the groups. If results reveal a “good model fit” researchers are justified to continue testing additional areas of ME/I (such as the equivalence of factor loadings and item intercepts); however, a “poor model fit” warrants no further testing (Vandenberg & Lance, 2000).

ME/I and the WISC-III

Maller and Ferron (1997) investigated the factor structure of the WISC-III for a sample of severely and profoundly deaf children ($N = 110$) and compared it to the factor structure of the WISC-III identified for the standardization sample ($N = 2200$). All WISC-III subtest directions were presented in the child’s primary mode of communication. The correlation matrix from the WISC-III standardization sample was unstandardized, resulting in a covariance matrix used in the analyses. First, the four-factor model identified for the standardization sample of the WISC-III (i.e., Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed) was tested on the sample of deaf children (i.e., test of configural invariance). The model had “good” fit both the standardization sample ($\chi^2(59) = 323.02$, goodness-of-fit [GFI] = .978, root mean square

residual [RMSR] = .261, Tucker-Lewis Index [TLI] = .970, comparative fit index [CFI] = .977) and the deaf sample (χ^2 (59) = 52.88, GFI = .935, RMSR = .625, NNFI = 1.013, CFI = 1.00). Next, the factor loadings were constrained to be equal across groups (i.e., test of metric invariance). Overall, metric invariance was not found. However, tests of partial invariance revealed that the unstandardized path coefficients were invariant across the two samples on all subtests except: Block Design, Arithmetic, Digit Span, and Coding. In sum, these four subtests did not indicate the latent construct of intelligence in a similar way across the two samples. Interestingly, the verbal subtests of the WISC-III had equal factor loadings among the hearing and deaf samples.

Purpose of the Present Study

No research to date has investigated the factor structure of intelligence of D/HOH children with respect to the CHC theory as measured with the WISC-IV. A number of research studies have found theoretical support of the CHC theory for hearing children and adults (e.g., Bickley, Keith & Wolfe, 1995; Carroll, 2003; McGrew & Woodcock, 2001; Roberts et al., 2000; Roid, 2003; Taub & McGrew, 2004), but similar research for D/HOH individuals is nonexistent. The research question of interest is, “Is the CHC theory, as measured with the WISC-IV, an appropriate model of intelligence for D/HOH children?” Using similar procedures outlined in the Maller and Ferron (1997) study discussed above, the purpose of the present study is to explore the structure of intelligence for D/HOH children within the context of the WISC-IV and compare it to the structure of the WISC-IV identified for the norm group. In other words, when compared to the norm group, is WISC-IV ME/I established for the D/HOH children? Because the WISC-IV is a current intelligence test

designed in accordance with the CHC theory, it will be used as the tool for assessing the structure of intelligence for this special population.

Documenting ME/I (or lack thereof) for the WISC-IV with D/HOH children and thus gaining insight into the structure of intelligence for this special group has important theoretical and practical implications. Theoretically, understanding the structure of intelligence for D/HOH children will help guide understanding of the nature and development of cognitive abilities and whether/how they are influenced by hearing impairment, and the development and use of appropriate cognitive tests for D/HOH examinees. Practically, investigating the measurement invariance of the WISC-IV with D/HOH children will provide important data for the validity (or lack thereof) of this scale with this exceptional population.

Hypotheses

Hypothesis One: Configural Invariance

I predict that the WISC-IV factor model for the D/HOH sample does not differ significantly from the WISC-IV factor model reported for the norm group. The null hypothesis states that the factor models for the two groups are statistically equal (i.e., invariant). If the results fail to reject the null hypothesis (i.e., the factor patterns did not differ across groups) and configural invariance is established, then I plan to proceed to Hypothesis Two to conduct a more stringent test of ME/I (i.e., metric invariance). If the null hypothesis is rejected and I find a lack of configural invariance, then I plan to conduct an EFA to find the best, most parsimonious WISC-IV factor model for my sample.

Hypothesis Two: Metric Invariance

I predict that the WISC-IV variable factor loadings for the D/HOH sample do not differ significantly from the same WISC-IV variable factor loadings for the norm group. A test of metric invariance will be conducted by constraining the matrix of factor loadings to be equal across the two groups. The null hypothesis states that the factor loadings of the two groups are statistically equal ($H_0: \Lambda^{\text{hearing}} = \Lambda^{\text{D/HOH}}$). If the results fail to reject the null hypothesis (i.e., the factor loadings do not differ across groups) and metric invariance is established, then I plan to proceed to Hypothesis Three. If the null hypothesis is rejected and I find a lack of metric invariance, then I plan to conduct tests of partial metric invariance to determine the source(s) of the variance.

Hypothesis Three: Differences in Mean Scores

I predict that the mean PRI and VCI scores for the D/HOH sample are significantly lower than the mean PRI and VCI scores reported for the norm group in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003), which are thought to reflect population values ($M = 100$). These hypotheses will be tested using a one-tailed, one-sample *t*-test.

I predict that the mean PSI score for the D/HOH sample does not differ significantly from the mean PSI score reported for the norm group in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003), which is thought to reflect the population value (i.e., $M = 100$). This hypotheses will be tested using a two-tailed, one-sample *t*-test.

Hypothesis Four: Subtest Intercorrelations

I predict that the correlations among WISC-IV subtests are positive and reliably greater than zero (i.e., $p > 0$). This hypothesis will be tested by calculating 95% confidence

intervals. If the lower bound of the interval was greater than zero, I will conclude the hypothesis is supported. For a path diagram of the planned study hypotheses, please see Figure 2 (next page).

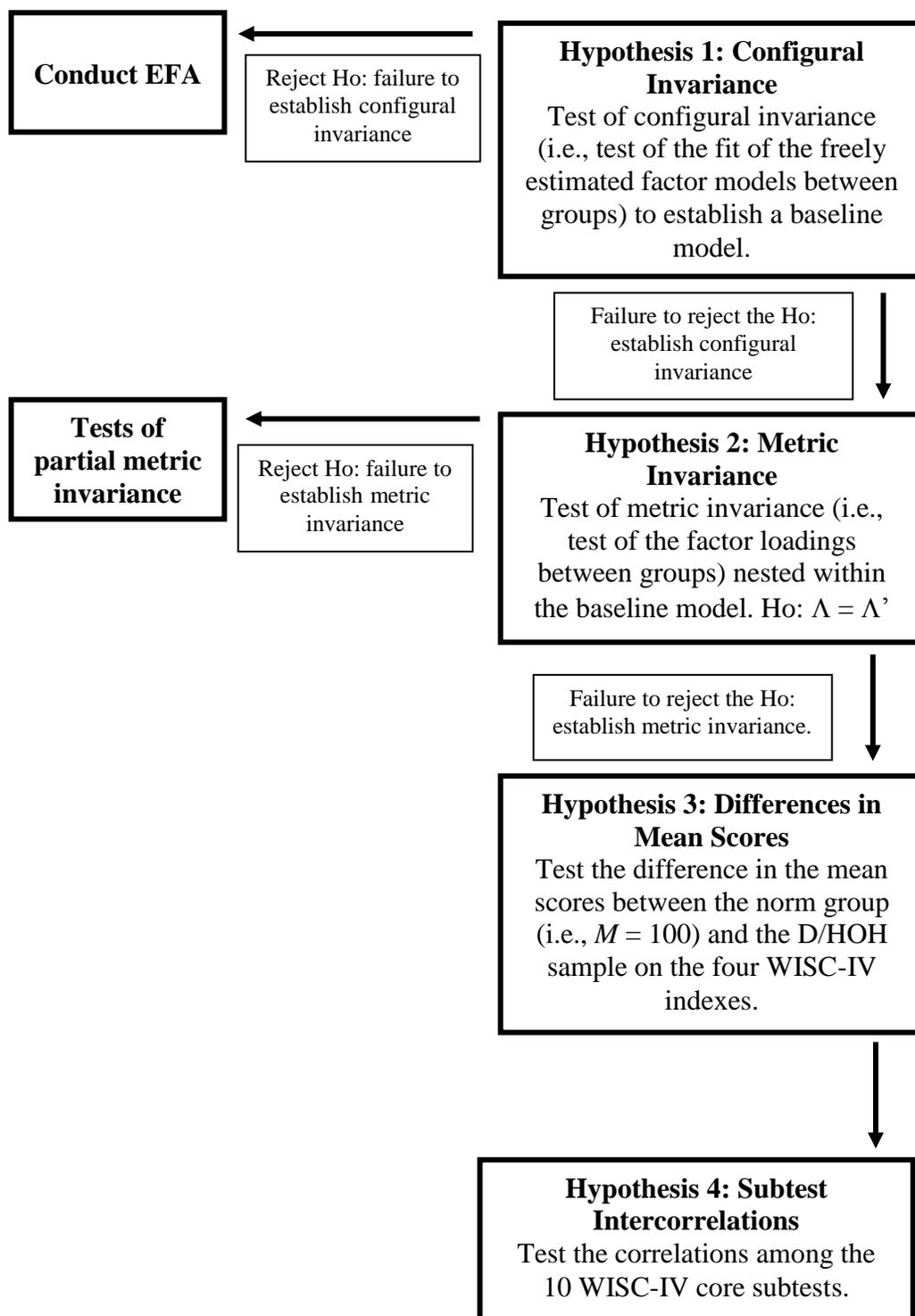


Figure 2. Path Diagram of Study Hypotheses.

CHAPTER 2

Method

Study Changes

Due to significant obstacles with data collection, the WISC-IV Working Memory (WM) subtests (i.e., Digit Span and Letter-Number Sequencing) scores and subsequently the Working Memory Index (WMI) scores were eliminated from all analyses. Therefore, the current study only examined a three-factor structure of the WISC-IV with deaf and hard-of-hearing (D/HOH) children as opposed to the four-factor structure described in *the WISC-IV Technical and Interpretive Manual* (Wechsler, 2003) for the norm group. The rationale for eliminating the WM subtests and Index score is provided below.

The WM subtests of the WISC-IV and D/HOH Examinees

It is common practice, outlined in the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 1999) that all individuals are tested in their primary mode of communication. Therefore, a psychologist wishing to give the WISC-IV to a D/HOH child who communicates via sign language must administer the assessment in the primary mode of communication-sign language. Administering the WISC-IV in sign language significantly changes the delivery structure and content of the Working Memory subtests. Because the subtests are administered vocally, and therefore tap the phonological loop (rather than the visual-spatial sketchpad asset of Working Memory), administration of these tests via visual signs or media alter the fundamental construct that is measured. For this reason, psychologists shy away from administering these subtests with this special population (N. Kordus, personal communication, May 21, 2011).²

There are two main reasons why psychologists do not administer the WM subtests of the WISC-IV to D/HOH children. The first reason is that when the subtest administration is changed from oral to visual (i.e., sign language), the construct being measured changes from auditory working memory to visual working memory. Research has shown that WM (i.e., *Gsm*) for deaf children is positively correlated with their oral language skills and negatively correlated with their degree of hearing loss (Marschark, 2006). Hearing individuals have longer memory spans than deaf individuals, and deaf individuals who use spoken language have longer memory spans than deaf individuals who use sign language (Mayberry & Eichen, 1991). The difference among these groups may be due to use of different encoding systems (i.e., phonological vs. visual-spatial), or to differences in working memory management. A plethora of evidence (e.g., MacSweeney, et al, 1996; Wilson & Emmorey, 1997a, 1997b) suggests that individuals who use spoken language rely on the phonological loop of working memory, whereas individuals who use sign language rely on the visual-spatial sketchpad (Marschark, 2003). Deaf and hearing individuals appear to have same working memory capacity, but signs take longer to articulate than words, and are less amenable to subvocalization strategies required for using phonological loop capacity, therefore negatively affecting the working memory capacity of deaf individuals (Marschark, 2003; 2006).

The second reason is that different WM rehearsal mechanisms (i.e., phonological loop vs. visual-spatial sketchpad) are based in different sensori-motor modalities, with unique processing constraints (e.g., spatial and temporal information; Wilson, et al., 1997). Wilson and colleagues (1997) found that deaf individuals performed similarly on forward and

backward recall of linguistic stimuli; in contrast, hearing individuals in their study (and in all norm groups) performed substantially worse on backward recall compared to forward recall. This finding suggests that WM for speech and sign differ in how they represent serial order information. Furthermore, deaf individuals were better than hearing individuals on nonlinguistic spatial memory, suggesting that expertise in visual-spatial language (i.e., sign language) may influence other forms of visual-spatial memory. Considering the differences in how the subtests are administered to D/HOH children and the different neurological processes evoked (compared to hearing children), psychologists question the content validity of the WM subtests with this special population and, therefore, do not administer these tests to D/HOH children.

Despite the empirical implications for administering the WM subtests of the WISC-IV to D/HOH children, there is also the practical reality that psychologists serving the D/HOH population do not administer the WM subtests of the WISC-IV to this special population. Only in very rare cases are the WM subtests administered and scored. Dr. Natasha Kordus, a psychologist at the California School of the Deaf, said that in the past five years, she has administered these subtests only about 10 times (personal communication, May 21, 2011). Other psychologists with whom the researcher has communicated during the data collection process have reported either eliminating the WM subtests from the WISC-IV core battery or choosing a different cognitive assessment (such as the Wechsler Nonverbal Scale of Intelligence, Wechsler & Naglieri, 2006).

In sum, there were insufficient WISC-IV WM data available to collect for the current study. Due to the empirical and practical reasons discussed above, the WM subtest and the

WMI scores have been dropped from the factor structure analyses. When a reference is made to the factor structure of the WISC-IV in the following discussion, the reader is reminded that the WISC-IV WM subtest (i.e., Digit Span and Letter-Number Sequencing) and WM Index scores are excluded along with the Full Scale IQ score.

Participants and Examinees

This study drew a distinction between participants and examinees. The participants were psychologists who volunteered to provide data on the WISC-IV with D/HOH children. The examinees were those D/HOH children on which the WISC-IV data were collected. In other words, the participants (i.e., the psychologists) tested the examinees (i.e., the D/HOH children) with the WISC-IV and supplied these data to me (i.e., the researcher).

Recruitment

The participants were practicing psychologists who were recruited at area conferences, postings on deafness/psychology-related Listservs and websites and by telephone (see Appendixes C and D). Interested psychologists provided their names and e-mail addresses (as e-mail served as the primary mode of communication for the duration of the study). These psychologists provided archival data on D/HOH examinees who were tested with the WISC-IV for purposes other than research (e.g., educational placement, clinical or vocational assessment).

Participants were directed to identify examinees for inclusion based on a set of criteria. The inclusion criteria were the following: (a) the examinee must be between 6 years, 0 months and 16 years, 11 months of age, (b) have a hearing loss (i.e., ranging from mild to profound) sufficiently significant to be identified as having a hearing disability, (c) have

prelingual onset of hearing impairment (i.e., defined in this study as a hearing loss that occurs prior to the age of 5), (d) have a hearing impairment as their primary disability (if they have more than one), and (e) have been tested on the WISC-IV as part of a previous psychological evaluation (e.g., educational placement, clinical diagnosis). Participants were requested to exclude examinees not meeting all the selection criteria.

Procedure

The participants were sent a consent form via e-mail, which they printed, signed, and returned by either ground mail or facsimile (see Appendix E). Due to the insecure nature of e-mail, this consent form also had a space designated for the participants to supply a password. This password was used to encrypt the files and aid in the maintenance of confidentiality throughout the duration of the study. Once the consent form was received, the participants were sent a pre-formatted Microsoft Excel spreadsheet to complete. In this spreadsheet, participants were asked to enter non-identifying examinee descriptive data and WISC-IV subtest scaled scores and index standard scores. This spreadsheet was encrypted; thus, the participants were required to enter their password to open the file. The spreadsheet was unique to each participant, meaning that the spreadsheet contained subject identification numbers specific to the participant and empty cells to insert the data. Participants were able to insert information on as many examinees as they had available to them. No participant's data were shared with other participants. Once participants entered the information into the spreadsheet, they returned the completed encrypted file to me via e-mail.

Demographic Data

I collected a variety of demographic information about each examinee. Tables 5 and 6 display the demographic data gathered for the examinees.

Table 5

General Background and Hearing Loss Data Gathered for Each D/HOH Child

Variables	Possible Values
Age	6 years, 0 months – 16 years, 11 months
Grade	1 st – 12 th grade
Gender	Male, Female
Ethnicity	White, Black, Hispanic, Asian/Pacific Islander, Other/Unknown
Additional disabilities	Any DSM-IV Diagnosis or disability identified in IDEIA-04
Primary Mode of Communication	ASL, Other Sign, Aural/Oral, Cued Speech, ASL and Oral, Other
Degree of Hearing Loss	Mild, Moderate, Severe, Profound, Unknown
Pure Tone Average	
Better ear with aid	All possible PTA ranges
Better ear without aid	All possible PTA ranges
Type of Hearing Loss	Sensorineural, Conductive, Mixed, Unknown
Age of Onset	Congenital, Prelingual, Unknown

Table 6

Setting and Test Administration Data Gathered for Each D/HOH Child

Variables	Possible Values
Setting	Residential School for the deaf, Commuter School for the deaf, Program within Public School, Public School with Supplemental Services, Private School, Public School with no services, Other
Reason for Referral	Initial Evaluation, Triennial Reevaluation, Change of Service/Diagnosis, Initial Evaluation for Admission, Other

WISC-IV Data

I also gathered WISC-IV subtest scaled scores and index standard scores for each examinee. Table 7 displays the WISC-IV data gathered for each examinee.

Table 7

WISC-IV Data Gathered for Each D/HOH Child

Subtest Scaled Scores	Standard Scores
Block Design	Verbal Comprehension Index
Similarities	Perceptual Reasoning Index
Picture Concepts	Processing Speed Index
Coding	
Vocabulary	
Matrix Reasoning	
Comprehension	
Symbol Search	

Predictions and Analyses

Sample Size

In addition to collecting WISC-IV data from practicing psychologists, I also used data that were previously collected as part of a research study entitled “The Reliability and Validity of WISC-IV Scores with Deaf and Hard-of-Hearing Children” (Krouse & Braden, 2011). The data collection procedures used for the previous project were identical to those of

the current study. Due to the methodological procedures of this study, the sample of examinees is considered a “convenience sample.” This study used sampling methods that drew upon existing frameworks (e.g., Listservs) rather than systematic, random sampling procedures. Therefore, the sample is a nonrandom group.

The current study has a total sample size of $n = 134$. Of this sample, 83 cases (62%) are considered to be “new” cases obtained from current data collection methods. A total of 51 cases (38%; that met the present study criteria) were extracted from the previous (i.e., Krouse & Braden, 2011) data sample. This yields a total combined sample size of 134. This sample size is considered to be sufficient for the proposed analyses and relatively consistent with the recommendations made by Meade and Bauer (2007).

Merging Samples

The 83 “new” cases and 51 “old” cases were merged for the study analyses. A two tailed ($p < .05$) Analysis of Variance (ANOVA) comparing the means of the VCI, PRI and PSI between to the two sample groups (i.e., “old data” vs. “new data”) was conducted. The results of the one-way ANOVA are presented in Table 8. There were no significant differences between the two sample groups for each of the WISC-IV Indexes. The two samples were merged and the study analyses were conducted using the combined sample ($n = 134$).

Table 8

ANOVA Source Table

Source	SS	df	MS	F
VCI Between Groups	143.53	1	143.53	0.443
VCI Within Groups	42801.10	132	324.25	<i>NS</i>
VCI Total	42944.63	133		
PRI Between Groups	7.25	1	7.25	0.036
PRI Within Groups	26256.45	132	198.91	<i>NS</i>
PRI Total	26263.70	133		
PSI Between Groups	548.46	1	548.46	2.22
PSI Within Groups	32617.93	132	247.11	<i>NS</i>
PSI Total	33166.39	133		

Converting Data

The WISC-IV correlation matrix for the norm group is available in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003). The correlation matrix for the norm group was converted into a variance-covariance matrix ($N=2200$). All raw WISC-IV data collected on the D/HOH sample was also converted into a variance-covariance matrix ($N=134$). The derived variance-covariance matrixes for both groups were used for all subsequent analyses.

Factor Model

I conducted the study analyses using the AMOS (Analysis of Moment Structures) 17.0 program. The first step for conducting the analyses was to specify the factor model (see Figure 3), which was used for all analyses. When conducting confirmatory factor analyses (CFA), the researcher must provide a scale for each latent factor. This is done one of two ways: (a) fixing the factor loading of one variable for each latent factor to a constant, usually 1.0, or (b) fixing the variance of each latent factor to a constant, usually 1.0 (Reise, Widaman, & Pugh, 1993). Most research investigating ME/I addresses this problem by fixing the factor loading of one of the variables for each latent factor to 1.0 (Millsap, 2011). Therefore, I fixed one of the variables for each of the latent factors to 1.0 (i.e., Block Design [BD], Similarities [SIM] and Coding [CD]).

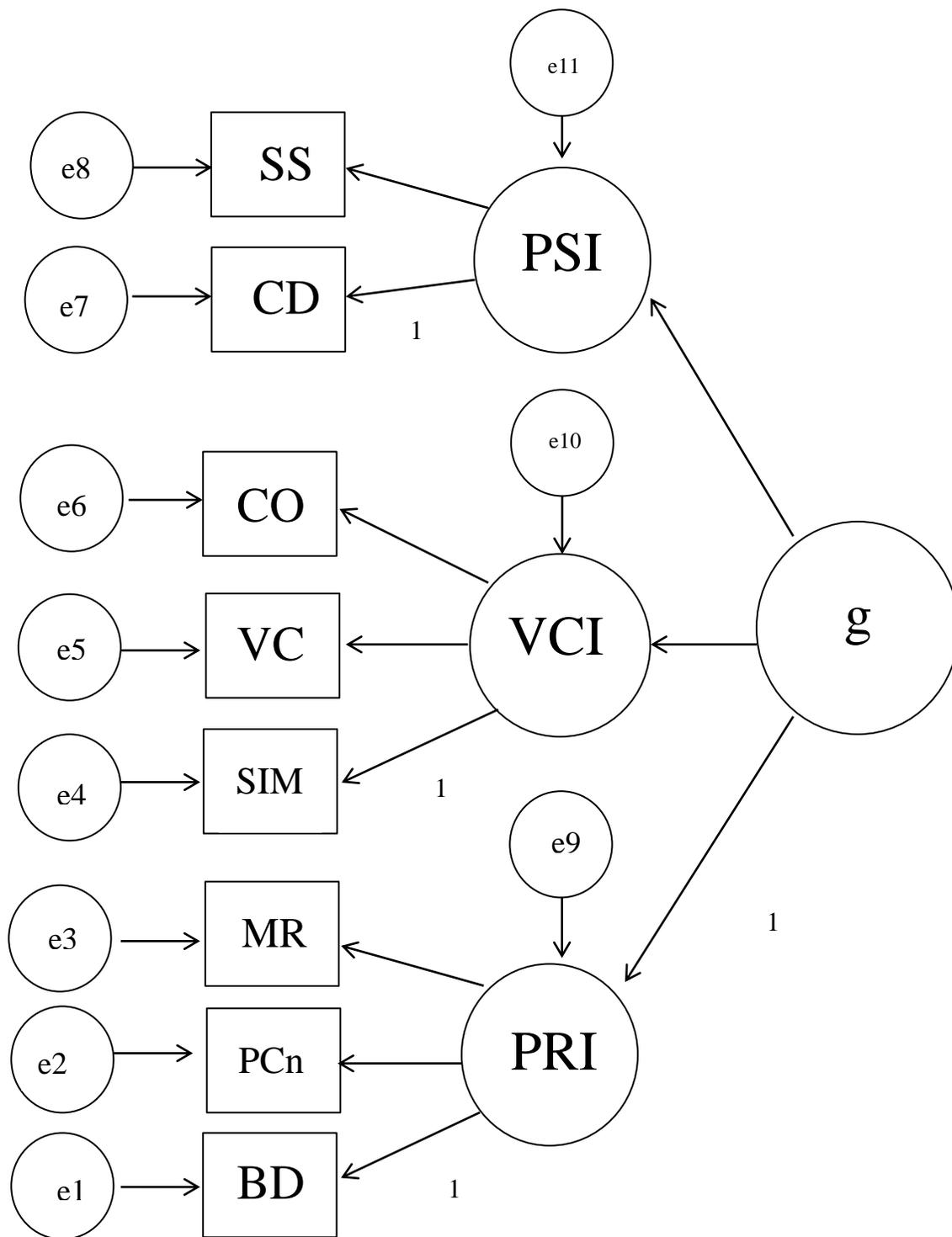


Figure 3. Factor Model of the WISC-IV. BD=Block Design, PCn=Picture Concepts, MR=Matrix Reasoning, SIM=Similarities, VC=Vocabulary, CO=Comprehension, CD=Coding, SS=Symbol Search, PRI=Perceptual Reasoning Index, VCI=Verbal Comprehension Index, PSI=Processing Speed Index, g=general intelligence, e=error.

Factor Details

Fit indices. In a multiple-group CFA context, the primary statistical measure of fit is the chi-square test, which provides an overall test of model fit to the data. However, the chi-square statistic is overly sensitive to trivial differences when the sample size is large (Bentler & Bonett, 1980). Therefore, additional ‘practical’ fit indices are commonly used to evaluate CFA models. Hu and Bentler (1998), and Reise, Widaman, and Pugh (1993), suggest researchers calculate two or more alternative fit indices (AFIs) when evaluating a model. Hu and Bentler (1999) recommend using the following additional fit indices: Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR).

The criteria to determine “good,” “adequate,” and “poor” fit are not clearly defined. Numerous researchers suggest different “rules of thumb” for determining good vs. poor model fit. For example, Keith (2005) suggests that for both the TLI and CFI values greater than .95 imply the model is a “good” fit for the data while values greater than .90 imply the model is an “adequate” fit. For the TLI and CFI, Keith suggests values less than .90 suggest “poor” fit, whereas AGFI values greater than .90 indicate a “good” model fit. RMSEA is a fit index in which smaller values suggest better model fit. For the RMSEA, values less than .06 suggest “good” fit and values between .06 and .08 suggest “adequate” fit. Similarly, smaller values for SRMR (i.e., < .08) suggest little difference between competing models (Keith, 2005).

MacCallum et al., (1996) recommended slightly different cutoff criteria. They believe that for TLI and CFI values greater than .90 suggest good model fit, and RMSEA values of

.05 or less reflect good fit whereas values between .05 and .08 are considered reasonable fit, and values ranging from .08 to .10 are considered mediocre fit. Arbuckle and Wothke (1999) suggest that, when assessing the difference chi-square/ df index, a score of two or less is considered most acceptable.

In an attempt to address the issues of cutoff criteria for the various AFIs, Hu and Bentler (1999) conducted a study assessing the ability of the AFIs to correctly reject the null hypothesis when it was false, and fail to reject it when it was true under a variety of conditions (e.g., different sample sizes, different model complexities). Hu and Bentler specifically investigated AFI pairs, pairing a fit index that is sensitive to differences between factor loadings in different models (e.g., TLI, CFI, RMSEA) with the SRMR, which is the most sensitive AFI to differences in the latent structures between models. In general, Hu and Bentler recommend a cutoff value of .95 for TLI, BL89, RNI, CFI or Gamma hat, and a cutoff rule of .05 or less for RMSEA and a cutoff of .06 or less for SRMR.

Taking into consideration the recommendations of previous research for evaluating overall model fit (e.g., Hu & Bentler, 1999; Meade, Johnson, & Braddy, 2008, Vandenberg & Lance, 2000) and the fit indexes reported in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003) for the norm group (i.e., χ^2 , $\Delta\chi^2$, χ^2/df , Adjusted Goodness-of-fit Index [AGFI], TLI, and RMSEA), I reported the following fit statistics: (a) chi-square, (b) chi-square/degrees of freedom (df) ratio, (c) TLI, (d) CFI, (e) RMSEA, and (f) SRMR. For AFI model fit cutoff criteria please see Table 9. Finally, all tests were conducted using an alpha level of .05.

Table 9

AFI Model Fit Cutoff Criteria

AFIs	Good Fit	Adequate Fit	Poor Fit
$\Delta\chi^2/dF$	< 2.0	--	> 2.0
TLI	> .95	.95 - .90	< .90
CFI	> .95	.95 - .90	< .90
RMSEA	< .05	.05 - .10	> .10
SRMR	< .06	.06 - .10	> .10

Note. AFI model fit cutoff criteria was gathered from the following sources: Keith (2005), MacCallum et al. (2006), Arbuckle & Wothke (1999) and Hu & Bentler (1999).

CHAPTER 3

Analyses and Results

Participants and Examinees

Nine participants from seven states across the nation (i.e., Arizona, California [2], Illinois, Kansas, Maryland [2], Massachusetts, and Pennsylvania) provided data on 134 D/HOH children (i.e., examinees). All examinees met the inclusion criteria. The examinees' ages ranged from 6.0 years to 16.83 years ($M = 12.16$, $SD = 2.92$). Demographic data were aggregated across age groups. Frequencies for gender, ethnicity, grade, current educational placement, reasons for referral/testing, primary mode of communication, degree of hearing impairment, type of hearing loss, age of onset, and comorbid diagnoses are reported in Table 10. Frequencies, means and standard deviations for age of onset (in months), degree of hearing loss (in dB), and PTA (in dB) with and without hearing aids for the better ear are reported in Table 11.

Table 10

Demographic Information

Demographic Categories	Sample Size (<i>n</i>)	Percentage (%)
Total <i>N</i>	134	100.0%
Gender		
Male	60	44.8%
Female	74	55.2%
Ethnicity		
White	64	47.8%
Black	7	5.2%
Hispanic	54	40.3%
Asian/Pacific Islander	6	4.5%
Unknown	3	2.2%
Grade		
First	7	5.2%
Second	9	6.7%
Third	13	9.7%
Fourth	14	10.4%
Fifth	10	7.5%
Sixth	14	10.4%

Table 10 (continued)

Demographic Information

Demographic Categories	Sample Size (<i>n</i>)	Percentage (%)
Grade (continued)		
Seventh	14	10.4%
Eighth	13	9.7%
Ninth	25	18.7%
Tenth	8	6.0%
Eleventh	6	4.5%
Missing	1	0.7%
Current Educational Placement		
Residential School for deaf students	66	49.3%
Commuter School for deaf students	42	31.3%
Program w/in Public School	9	6.7%
Public School w/ supplemental services	10	7.5%
Private School	3	2.2%
Public School w/ no services	3	2.2%
Other	1	0.7%

Table 10 (continued)

Demographic Information

Demographic Categories	Sample Size (<i>n</i>)	Percentage (%)
Reason for Referral/Testing		
Initial Evaluation for placement	38	28.4%
Triennial Reevaluation	65	48.5%
Change of service/diagnosis	5	3.7%
Initial Evaluation for Admission	1	0.7%
Other	25	18.7%
Primary Mode of Communication		
American Sign Language (ASL)	94	70.1%
Other Sign	8	6.0%
Oral/Aural	29	21.6%
ASL & Oral/Aural	2	1.5%
Cued Speech	1	0.7%
Degree of Hearing Impairment		
Mild (21-40 dB)	8	6.0%
Moderate (41-70 dB)	22	16.4%
Severe (71-90 dB)	28	20.9%
Profound (> 91 dB)	74	55.2%
Missing	2	1.5%

Table 10 (continued)

Demographic Information

Demographic Categories	Sample Size (<i>n</i>)	Percentage (%)
Type of Hearing Loss		
Conductive	1	0.7%
Sensorineural	123	91.8%
Mixed	4	3.0%
Unknown	6	4.5%
Age at Onset		
Congenital (at birth)	56	41.8%
Prelingual (before 2 years)	53	39.6%
Unknown	25	18.6%
Comorbid Diagnoses		
None	12	9.0%
ADHD	8	6.0%
ADHD, NVLD	1	0.7%
ADHD, ESL	1	0.7%
ADHD, LD	1	0.7%
ADHD, Adjustment Disorder	1	0.7%
ADHD, Spina Bifida Occulta	25	18.7%
ADHD, Expressive Language Disorder	1	0.7%

Table 10 (continued)

Demographic Information

Demographic Categories	Sample Size (<i>n</i>)	Percentage (%)
Comorbid Diagnoses (continued)		
ADHD, Adjustment Disorder with Anxiety, LD	1	0.7%
Language Based LD	1	0.7%
LD	5	3.7%
OHI	1	0.7%
OCD, NVLD	1	0.7%
Academically Gifted	1	0.7%
Tourette's Syndrome	1	0.7%
Duane's Syndrome	1	0.7%
Cleft Palate/Lip Double	1	0.7%
GAD	1	0.7%
ESL	1	0.7%
PDD	1	0.7%
Other	1	0.7%
Total Comorbid	67	50.0%

Note. Percentages may not equal 100% due to rounding. ADHD = Attention Deficit Hyperactivity Disorder, ESL= English as a Second Language, GAD = Generalized Anxiety Disorder, LD = Learning Disability, NVLD= Nonverbal Learning Disability, OCD= Obsessive Compulsive Disorder, OHI = Other Health Impaired, PDD = Pervasive Developmental Disorder.

Table 11

Demographic Information describing Examinees' Hearing Impairment

Characteristics of Hearing Impairment	<i>n</i>	<i>M</i>	<i>SD</i>
Age of Onset (in months)	100	10.96	14.17
Degree of Hearing Loss (in dB)	21	61.19	31.04
PTA of better ear without hearing aids (in dB)	96	82.82	23.21
PTA of better ear with hearing aids (in dB)	31	38.97	18.29

Note. PTA = Pure Tone Average; dB = decibels.

Descriptive Analyses

The mean and standard deviation of each of the WISC-IV Subtests and the Composite Scales are reported in Table 12.

Table 12

The Mean and Standard Deviation of the WISC-IV Subtests and Composite Scales

WISC-IV Subtests and Composite Scales	<i>M</i>	<i>SD</i>
Core Subtests		
Block Design(BD)	9.27	2.80
Similarities(SIM)	7.54	3.24
Picture Concepts(PCn)	9.34	2.87
Coding(CD)	8.67	2.48
Vocabulary(VC)	4.75	3.10
Matrix Reasoning(MR)	9.49	2.98
Comprehension(CO)	7.63	4.44
Symbol Search(SS)	8.92	2.94
Composite Scales		
Verbal Comprehension Index(VCI)	80.05	17.97
Perceptual Reasoning Index(PRI)	96.18	14.05
Processing Speed Index(PSI)	94.16	15.79

Note. The Core and Supplemental subtests have a mean of 10 and a standard deviation of 3; The Composite scales have a mean of 100 and a standard deviation of 15.

Preliminary Analyses: Restriction of Range

A correlation coefficient could be spuriously low if it is based on a sample of homogeneous subjects for whom there is a restriction of range on the characteristic being measured. To ensure that subtest intercorrelation coefficients are not unduly influenced by an unusual range of scores, I conducted a one-tailed, one sample F test for equal variances assuming a population variance of 9 (i.e., $SD = 3$, $SD^2 = 9$) for WISC-IV subtest scaled scores ($p < .05$). The F test for equal variances was only calculated for those subtests scores for which the variance was less than the population variance (i.e., cases in which restriction of range could be a problem).

I calculated an F test of equal variance for the Block Design ($SD = 2.80$), Picture Concepts ($SD=2.87$), Matrix Reasoning ($SD=2.98$), Coding ($SD=2.48$) and Symbol Search ($SD=2.94$) subtests. The F -ratio was significant for the Coding subtest, indicating a restriction of range, $F(2199, 133) = 1.46$, $p < .05$. Therefore, I reported corrected subtest intercorrelations that included the Coding subtest, using the common restriction of range correction formula for bivariate correlations (Guilford, 1965; see Appendix F).

F -ratios for the Block Design subtest, $F(2199, 133) = 1.15$, $p > .05$, Picture Concepts subtest $F(2199, 133) = 1.09$, $p > .05$, Matrix Reasoning $F(2199, 133) = 1.01$, $p > .05$ and Symbol Search subtest $F(2199, 133) = 1.04$, $p > .05$, indicated restriction of range was not likely to influence correlations between these subtests and other variables. Therefore, I did not correct correlations among these subtests (and report uncorrected values only).

Hypotheses

Hypothesis One: Configural invariance. I predicted that the WISC-IV factor model for the D/HOH sample ($N=134$) will not differ significantly from the WISC-IV three-factor model reported for the norm group ($N=2200$). The data sets from the two groups (i.e., norm group and D/HOH sample) were analyzed simultaneously in AMOS 17.0, allowing item parameters (i.e., factor loadings) to be freely estimated across the groups (i.e., factor loadings were free to vary across groups with no equality constraints imposed). Figures 4 and 5 show the standardized estimates for the norm group and D/HOH sample, respectively.

The goodness-of-fit values for the configural model were $\chi^2(df) = 155.71(34)$, $\chi^2/df = 4.58$, $p < .000$, TLI = .974, CFI = .984, RMSEA = .039, SRMR = .021. The χ^2 fit index suggests lack of configural invariance; however, this index is extremely sensitive to large sample sizes (Bentler & Bonett, 1980). The other fit indexes, TLI, CFI, RMSEA and SRMR values indicated good fit for the configural model. Overall, the data for both groups (norm group and D/HOH sample) adequately fit the WISC-IV three-factor model illustrated in Figure 3. The factor structure for the two groups was invariant. This model served as the baseline model against which more stringent tests of invariance (e.g., metric invariance and scalar invariance) were nested and compared.

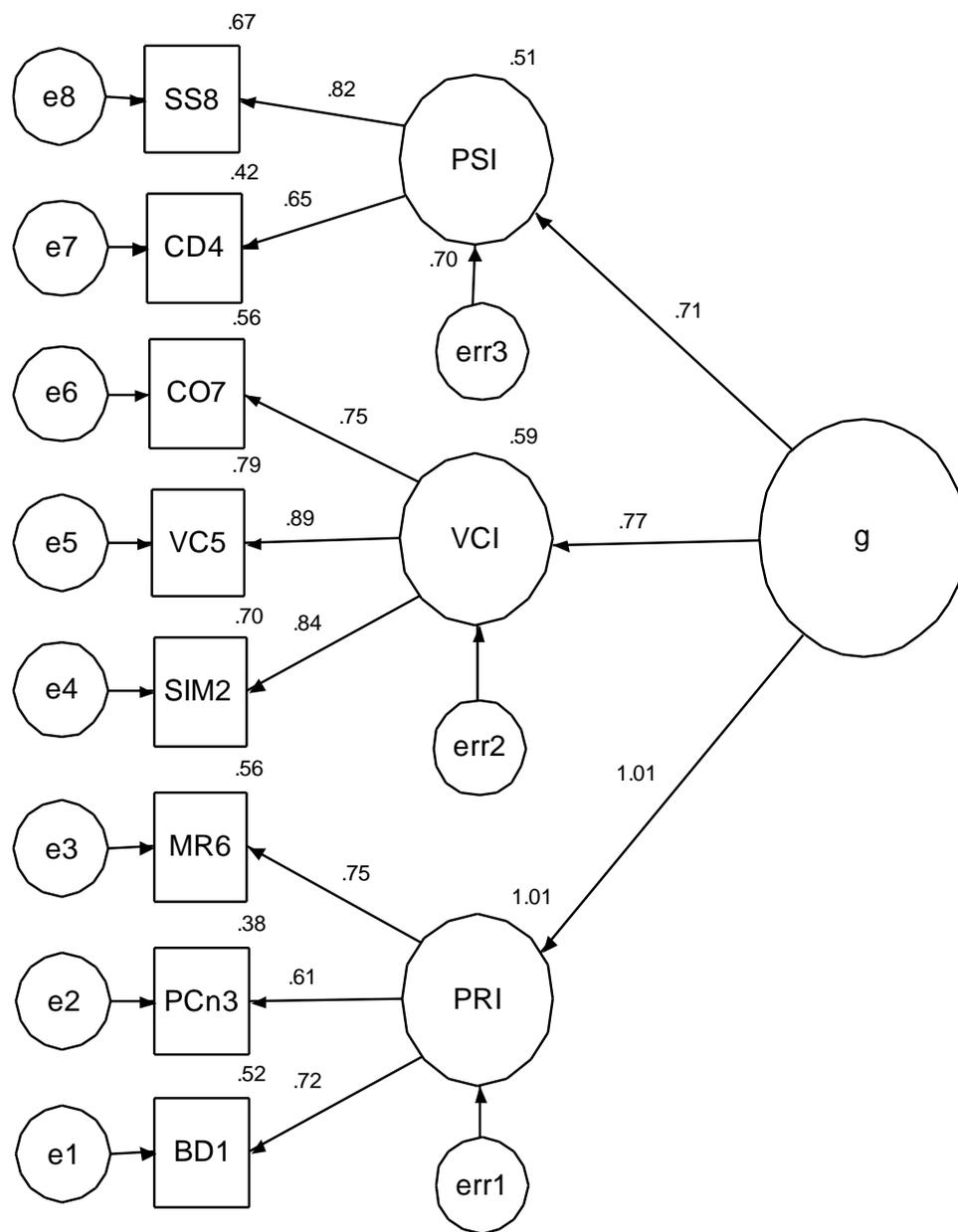


Figure 4. Standardized estimates for the norm group.

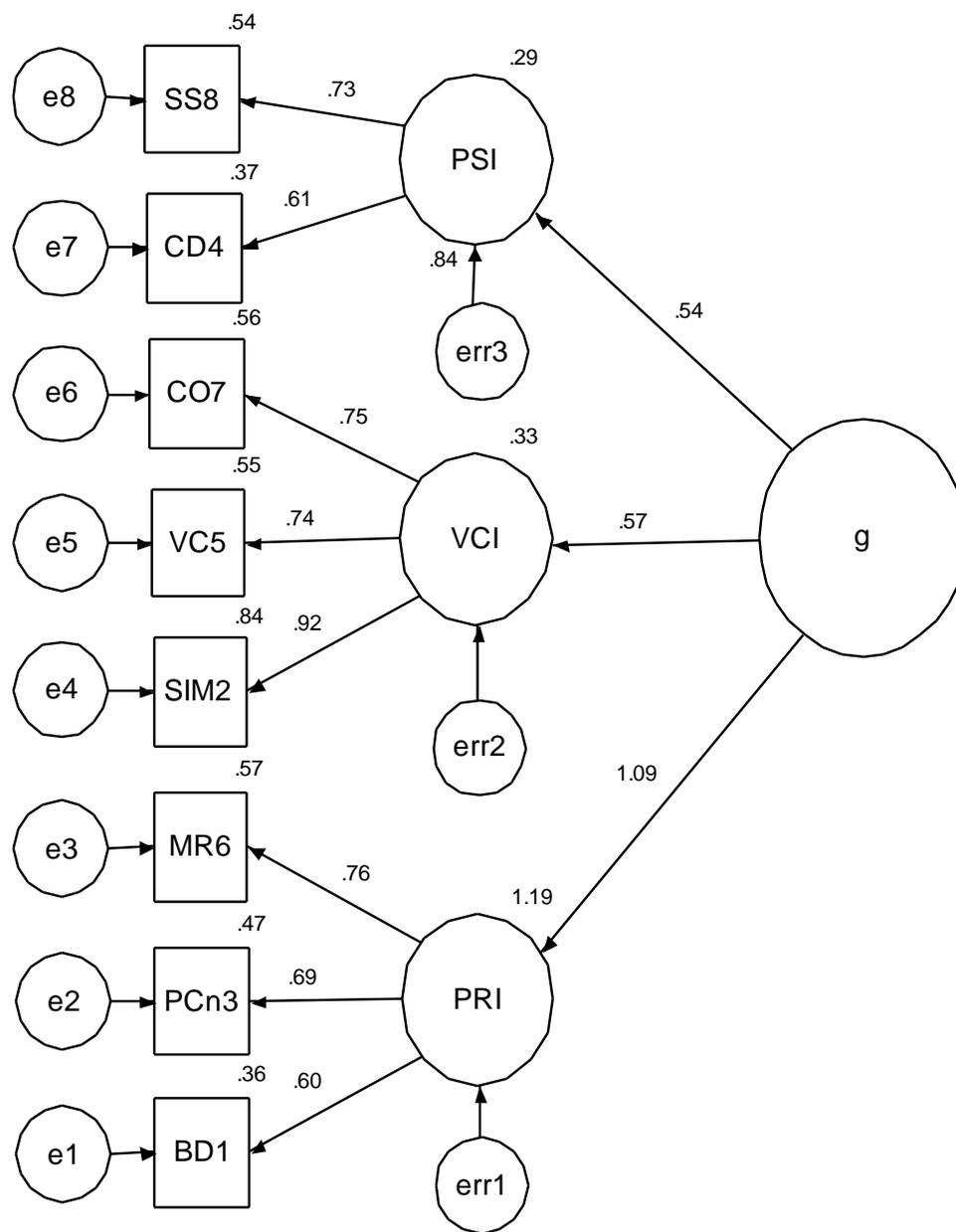


Figure 5. Standardized estimates for the D/HOH sample.

Hypothesis two: Metric invariance. I predicted that the WISC-IV variable factor loadings for the D/HOH sample would not differ from the same WISC-IV variable factor loadings for the norm group. I conducted a test of metric invariance by constraining all factor loadings to be equal across the two groups. I nested the metric invariance model within the baseline model, and used the likelihood ratio test (i.e., $\Delta \chi^2$) to compare models. The $\Delta \chi^2$ was significant, $\Delta \chi^2(\Delta df) = 24.92(5), p < .000$, which means that constraining the factor loadings to be equal across groups caused a significant change in the model fit. Metric invariance was not established between the two groups.

Next, I conducted tests of partial metric invariance to determine the source(s) of the lack of invariance (Byrne et al., 1989). To do this, I constrained the factor loadings for each of the composites separately (i.e., factor loadings that comprise the PRI, VCI and PSI) and allowed all other factor loadings to be free across the two groups. I then compared the constrained models to the baseline model one at a time.

In the first test, I constrained only the factor loadings that comprise the PRI to be equal across the two groups (i.e., Block Design [BD], Picture Concepts [PCn] and Matrix Reasoning [MR]). The $\Delta \chi^2$ was not significant, $\Delta \chi^2(\Delta df) = 3.89(2), p > .05$, which supports metric invariance across the two groups for the PRI subtests.

For the second test, I constrained only the factor loadings that comprise the VCI to be equal across the two groups (i.e., Similarities [SIM], Vocabulary [VC] and Comprehension

[CO]). The $\Delta \chi^2$ was significant, $\Delta \chi^2(\Delta df) = 20.63(2)$, $p < .001$, which does not support metric invariance across the two groups for the VCI subtests.

For the third test, I constrained only the factor loadings that comprise the PSI to be equal across the two groups (i.e., Coding [CD] and Symbol Search [SS]). The $\Delta \chi^2$ was not significant, $\Delta \chi^2(\Delta df) = 0.22(1)$, $p > .05$, which supports metric invariance across the two groups for the PSI subtests.

Additional tests of invariance. Although I did not originally propose additional tests of factor invariance, I decided to conduct a more stringent test of factor invariance (i.e., a test of scalar invariance) for the PRI and PSI subtests due to the fact that metric invariance was established for PRI and PSI composites (separately).

Some researchers (e.g., Raju, Laffitte & Byrne, 2002; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000) claim that to make valid group comparisons of mean scores, scalar invariance should first be established. Scalar invariance investigates the equivalence of item intercepts across groups. Item intercepts are the values of the observed scores when the latent trait is zero, akin to regression intercepts. Observed mean differences at the item level should reflect the factor mean differences (Vandenberg & Lance, 2000). Scalar invariance has been interpreted by some researchers as a test of systematic response bias that differs between groups. In this case, differences in intercepts may reflect undesirable biases for one group compared to the other (Vandenberg & Lance). Unless scalar invariance is supported, the validity of inferences pertaining to cross-group comparisons is questionable.

I assessed scalar invariance for the PRI and PSI composites separately in continuation of the tests of partial invariance. For the test of scalar invariance, I constrained the item intercepts to be equal across groups for each composite (i.e., PRI and PSI separately). I nested the scalar invariance model within the metric invariance model (for PRI and PSI, respectively) and observed the $\Delta \chi^2$.

In the first test, I constrained the PRI subtest intercepts (i.e., Block Design [BD], Picture Concepts [PCn] and Matrix Reasoning [MR]), in addition to the factor loadings, to be equal across the two groups. The $\Delta \chi^2$ was significant, $\Delta \chi^2 (\Delta df) = 10.07(3), p < .05$, which does not support scalar invariance across the two groups for the PRI subtests. Due to the fact that scalar invariance was not established, I conducted no further invariance testing for the PRI composite.

In the second test, I constrained the PSI subtest intercepts (i.e., Coding [CD] and Symbol Search [SS]), in addition to the factor loadings, to be equal across the two groups. The $\Delta \chi^2$ was significant, $\Delta \chi^2 (\Delta df) = 33.96(2), p < .05$, which does not support scalar invariance across the two groups for the PSI subtests. Due to the fact that scalar invariance was not established, I conducted no further invariance testing for the PSI composite. Unconstrained scalar values for the D/HOH sample and Norm group are reported in Table 13 along with 95% confidence intervals and the effect size of the intercept difference between groups. For a summary of all invariance test results, the reader is referred to Table 14.

Table 13

Scalar Invariance Intercept Values

Unconstrained Scalar Values			
Subtest	D/HOH Sample	Norm Group	Cohen's <i>d</i>
Block Design (BD)	9.27	10.0	0.24
[95% C.I.]	[8.80, 9.74]	[9.88, 10.13]	
Picture Completion (PCn)	9.34	10.0	0.22
[95% C.I.]	[8.85, 9.83]	[9.88, 10.13]	
Matrix Reasoning (MR)	9.49	10.0	0.17
[95% C.I.]	[8.99, 9.99]	[9.88, 10.13]	
Coding (CD)	8.67	10.0	0.45
[95% C.I.]	[8.25, 9.09]	[9.88, 10.13]	
Symbol Search (SS)	8.92	10.0	0.36
[95% C.I.]	[8.42, 9.42]	[9.88, 10.13]	

Table 14

Results from Multiple Tests of Invariance

Test	Model	$\chi^2(df)$	p	$\Delta \chi^2 (\Delta df)$	p
Whole Model	Configural	155.71(34)	.000	-----	-----
	Metric	180.64(39)	.000	24.92(5)	.00
Partial: PRI	Metric	159.60(36)	.000	3.89(2)	NS
	Scalar	169.67(39)	.000	10.07(3)	.02
Partial: VCI	Metric	176.35(36)	.000	20.64(2)	.00
Partial: PSI	Metric	155.94(35)	.000	0.22(1)	NS
	Scalar	189.90(37)	.000	33.96(2)	.00

Hypothesis three: Differences in mean scores. I predicted that the mean VCI for the D/HOH sample would be lower than the mean VCI reported for the norm group in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003). I used a one-tailed, one-sample t -test with an alpha level of $p = .05$ to test the difference between the sample and norm group (i.e., population) means. This hypothesis was supported. The mean VCI for the D/HOH sample ($M = 80.05$, $SD = 17.97$) was lower than the population mean ($M = 100$, $SD = 15$; $t(133) = -12.85$, $p < .001$). I also calculated Cohen's d to determine the effect size of this difference ($d = 1.31$).

I predicted the mean PRI for the D/HOH sample would be lower than the mean PRI reported for the norm group in the *WISC-IV Technical and Interpretive Manual* (Wechsler,

2003). I used a one-tailed, one-sample t -test with an alpha level of $p = .05$ to test the difference between the sample and norm group (i.e., population) means. This hypothesis was supported. The mean PRI for the D/HOH sample ($M = 96.18$, $SD = 14.05$) was lower than the population mean ($M = 100$, $SD = 15$; $t(133) = -3.15$, $p < .05$). I also calculated Cohen's d to determine the effect size of this difference ($d = 0.26$).

I predicted the mean PSI for the D/HOH sample would *not* differ from the mean PSI reported for the norm group in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003). I used a two-tailed, one-sample t -test with an alpha level of $p = .05$ to test the difference between the sample and norm group (i.e., population) means. This hypothesis was not supported. The mean PSI for the D/HOH sample ($M = 94.16$, $SD = 15.79$) was lower than the population mean ($M = 100$, $SD = 15$; $t(133) = -4.28$, $p < .001$). I also calculated Cohen's d to determine the effect size of this difference ($d = 0.39$).

Hypothesis four: Subtest intercorrelations. I predicted that the correlations among WISC-IV subtests for the D/HOH sample would be positive and reliably greater than zero ($\rho > 0$) at alpha $p = .05$. Pearson product-moment correlations were calculated between the WISC-IV subtest scaled scores. Next, I calculated a 95% confidence interval for each correlation. If the lower bounds of the confidence intervals were greater than zero, I would conclude that my prediction was supported. However, if the confidence intervals included zero, the hypothesis would not be supported. The correlations for the Coding subtest were corrected for restriction of range. The confidence intervals around the corrected correlations were analyzed in hypothesis four.

I calculated twenty-eight Pearson product-moment correlations among the eight WISC-IV subtests analyzed in this study. Of the 28 sets of correlation confidence intervals, 3 contain zero. Therefore, this hypothesis was partially supported. See Table 15 for the correlation matrix.

Table 15

Correlations among WISC-IV Subtests (with 95% Confidence Intervals reported)

Subtests		PCn	MR	SIM	VC	CO	SS	CD ^a	CD ^b
BD	<i>r</i>	.39	.50	.30	.06	.19	.31	.25^a	.30^b
	[95% C. I.]	[.24, .52]	[.36, .62]	[.14, .45]	[-.11, .23]	[.02, .35]	[.15, .46]	[.08, .40]	[.14, .45]
PCn	<i>r</i>	--	.50	.49	.36	.32	.30	.20^a	.24^b
	[95% C. I.]	--	[.36, .62]	[.35, .61]	[.20, .50]	[.16, .46]	[.13, .44]	[.03, .35]	[.07, .39]
MR	<i>r</i>	--	--	.30	.29	.36	.30	.29^a	.34^b
	[95% C.I.]	--	--	[.13, .44]	[.13, .44]	[.21, .50]	[.13, .44]	[.13, .44]	[.18, .48]
SIM	<i>r</i>	--	--	--	.68	.68	.21	.17 ^a	.20^b
	[95% C. I.]	--	--	--	[.58, .76]	[.58, .76]	[.04, .36]	[.00, .33]	[.03, .36]
VC	<i>r</i>	--	--	--	--	.60	.15	-.05 ^a	-.06 ^b
	[95% C. I.]	--	--	--	--	[.48, .70]	[-.02, .31]	[-.22, .12]	[-.23, .11]
CO	<i>r</i>	--	--	--	--	--	.24	.24^a	.28^b
	[95% C.I.]	--	--	--	--	--	[.07, .39]	[.08, .40]	[.12, .43]
SS	<i>r</i>	--	--	--	--	--	--	.45^a	.50^b
	[95% C.I.]	--	--	--	--	--	--	[.30, .56]	[.36, .65]

Note. BD=Block Design; PCn=Picture Concepts, MR=Matrix Reasoning, SIM=Similarities, VC=Vocabulary, CO=Comprehension, SS=Symbol Search, CD=Coding, C.I. = Confidence Interval; Significant correlations are in bold font; ^a correlations *not* corrected for restriction of range; ^b correlations corrected for restriction of range.

Chapter 4

Discussion

The purpose of the present study was to explore the structure of intelligence for D/HOH children within the context of the WISC-IV and compare it to the structure of the WISC-IV identified for the norm group. I did so by examining the various levels of measurement invariance (i.e., configural, metric, and scalar) of the WISC-IV factor model across the two groups (i.e., norm group and the D/HOH sample). It is important to remember that data were unavailable for the Working Memory Index (WMI) subtests (i.e., Digit Span and Letter Number Sequencing) and the WMI composite and, therefore, were excluded from all analyses. As a result, only a three-factor WISC-IV model was assessed (as illustrated in Figure 3).

I could not find any research that has investigated the factor structure of intelligence of D/HOH children with respect to the CHC theory as measured with the WISC-IV. Documenting configural, metric and/or scalar invariance (or lack thereof) for the WISC-IV with D/HOH children has important theoretical and practical implications. Theoretically, understanding the structure of intelligence for D/HOH children will help guide understanding of the nature and development of cognitive abilities, and whether/how they are influenced by hearing impairment. These data will also help guide the development and use of appropriate cognitive tests for D/HOH examinees. Practically, investigating the measurement invariance of the WISC-IV with D/HOH children will provide important data for the validity (or lack thereof) of this scale with this exceptional population.

Hypothesis One: Configural Invariance

Hypothesis one assessed the configural invariance between the norm group and D/HOH sample on the three-factor model illustrated in Figure 3. Overall, the data supported the first hypothesis. The χ^2 fit index suggested lack of configural invariance; however, this index is extremely sensitive to large sample sizes (Bentler & Bonett, 1980). The other fit indexes, TLI, CFI, RMSEA and SRMR indicated good fit for the configural model. Therefore, I concluded that the data for both groups (norm group and D/HOH sample) adequately fit the WISC-IV three-factor model illustrated in Figure 3 and that the factor structure for the two groups is invariant.

The support for a three-factor model (similar to the norm group) differs from the previous factor analyses reported for the Wechsler Scales. Sullivan and Schulte (1992) analyzed the factor structure of the WISC-R with D/HOH children ($N = 368$) using an exploratory factor analysis procedure. The factor structure of the WISC-R subscales (except Mazes) was analyzed using a principal-axis factoring method followed by a varimax rotation of the factor solution. Factor loadings greater than .50 were required for assignment of subscales to factors. The authors reported that only two, not three, factors (i.e., Language Comprehension and Visual-Spatial Organization) described the WISC-R data for their D/HOH sample. The Language Comprehension and Visual-Spatial Organization factors respectively corresponded to Verbal and Performance scales; the Freedom from Distractibility (similar to Working Memory) did not emerge in their D/HOH sample. Sullivan found similar results in another study (Sullivan & Montoya, 1997) of the WISC-III with D/HOH children using the same exploratory factor analysis procedures. Sullivan identified

the same two factors (i.e., Language Comprehension and Visual-Spatial Organization) for the D/HOH sample. Again, the Freedom of Distractibility and Processing Speed factors were not found for their D/HOH sample.

Interestingly, Maller and Ferron (1997) investigated the configural invariance of the WISC-III between a sample of D/HOH children ($N = 110$) and the norm group using a multi-sample confirmatory factor analysis procedure. The data from both groups adequately fit the WISC-III four-factor model. Maller and Ferron found the same four factors (i.e., Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed) for both groups. This finding differed from the previous studies (discussed above) and is largely consistent with my current findings. There are at least two reasons why my results may be similar to those of Maller and Ferron (1997) and differ from the studies of the Wechsler Scales conducted by Sullivan and Schulte (1992) and Sullivan and Montoya (1997).

First, the WISC-III was the first version of the WISC to adopt a four-factor model structure (although scores were still primarily interpreted using the VIQ/PIQ dichotomy). This four-factor structure is more consistent with the CHC theory than the previous two-factor structure. The WISC-IV completely abandoned the VIQ/PIQ dichotomy and replaced it with the four-factor structure as the *primary* level of score interpretation (Wechsler, 2003). It is possible that the WISC-IV has been better designed, in part because it was intended to be more closely aligned with the CHC theory than previous editions, and in part because of better, newer test construction techniques and changes in content. Therefore, it is possible that I found the same factor structure for the D/HOH sample and the norm group because the

more recent versions are better designed and better aligned with CHC theory than the WISC-R.

Second, it is possible I found corroborating results with Maller and Ferron (1997) and conflicting results from the studies by Sullivan and Schulte (1992) and Sullivan and Montoya (1997) due to differing statistical procedures. The studies conducted by Sullivan and Schulte and Sullivan and Montoya used exploratory factor analysis (EFA) procedures whereas Maller and Ferron and I used confirmatory factor analysis (CFA) procedures.

There are several differences between EFA and CFA procedures (such as testing measurement invariance) that may attribute to differences in study findings. With EFA procedures, the researcher holds no specific expectations as to the nature and number of factors that will be extracted from the data. In contrast, CFA procedures allow the researcher to test a specific model fit (Thompson, 2004). EFA procedures require the researcher to choose a factor rotation that either requires all factors to be correlated or uncorrelated. However, with CFA, the researcher can, in the same model, allow some factors to be correlated and others to be uncorrelated (Thompson, 2004). In addition, EFA allows all variables to freely load onto all factors, which can result in multiple cross loadings (although factor rotation is used to minimize the magnitude of cross loadings). With CFA, the simple structure is obtained by specifying which variables load onto which factor, usually producing more parsimonious results than EFA approaches (Brown, 2006). Furthermore, with EFA all parameters in the factor model are estimated simultaneously. Equality constraints are not possible with EFA approaches. However, with CFA, the researcher(s) can constrain some parameters (e.g., factor loadings) to be equal across groups while having others vary freely

between groups (Thompson, 2004). “An important capability of CFA that is more difficult to realize in EFA is the possibility of testing a hypothesized factor structure in two separate samples simultaneously” (Comrey & Lee, 1992, p. 316). Therefore, the EFA procedures do not provide quantitative estimates of model fit between two groups. These differences in data analysis procedures may attribute to the different findings among studies. However, because changes in test content and data analysis methods are confounded between my study and those by Sullivan and colleagues, I cannot draw strong conclusions regarding the cause of the differences.

Hypothesis Two: Metric Invariance

With configural invariance established, the next step was to assess the metric invariance between the norm group and D/HOH sample. The goal of the second hypothesis was to assess whether the observed variables (i.e., WISC-IV subtests) were related to the three WISC-IV factors (i.e., CHC “second order” abilities) in the same way, and to the same degree, for D/HOH children as compared to the norm group. I predicted that the WISC-IV variable factor loadings for the D/HOH sample would *not* differ from the same WISC-IV variable factor loadings for the norm group. My hypothesis was not supported. In general, the factor loadings differed across the two groups. However, because the entire three-factor model was tested, I could not determine which loadings differed and which loadings (if any) were invariant across the two groups. Therefore, I conducted tests of partial metric invariance to determine the source(s) of the lack of invariance (i.e., specific WISC-IV subtests with different factor loadings between the groups).

Test of partial metric invariance: PRI. In the first test of partial metric invariance, only the factor loadings that comprise the PRI were set to be equal across the two groups (i.e., Block Design [BD], Picture Concepts [PCn] and Matrix Reasoning [MR]). Metric invariance was established across the two groups for the PRI subtests. The conclusion that the PRI subtest factor loadings are invariant across the norm group and D/HOH sample are largely consistent with previous research findings (e.g., Maller & Ferron, 1997). Maller and Ferron (1997) investigated the factor structure of the WISC-III for a sample of severely and profoundly deaf children ($N = 110$) and compared it to the factor structure of the WISC-III identified for the norm group ($N = 2200$). Tests of partial invariance revealed that the unstandardized path coefficients were invariant across the two samples on all Perceptual Organization Index subtests except Block Design. The Block Design subtest had a higher factor loading for the D/HOH sample (.888) compared to the norm group (.795).

The subtests of the PRI are considered the “nonverbal” or language-reduced subtests of the WISC-IV. The Block Design subtest measures Visual Processing (G_v) and the Matrix Reasoning and Picture Completion subtests measure Fluid Intelligence (G_f). In general, language-reduced measures of intelligence are recommended for assessing the cognitive abilities of D/HOH children (Braden & Athanasiou, 2005; Bradley –Johnson & Evans, 1991; Sullivan & Vernon, 1979; Zieziula, 1982). The current finding that the PRI subtest factor loadings are invariant between the norm group and D/HOH sample provide evidence for the validity (based on internal structure) of these subtests for use with D/HOH children as a measure of cognitive ability, and also suggests that cognitive processes associated with perceptual reasoning are similarly structured within children regardless of hearing status.

Test of partial metric invariance: VCI. In the second test of partial metric invariance, I constrained only the factor loadings that comprise the VCI to be equal across the two groups (i.e., Similarities [SIM], Vocabulary [VC] and Comprehension [CO]). Metric invariance was *not* established across the two groups for the VCI subtests. Interestingly, the Comprehension subtest loaded onto the second-order, latent factor VCI (i.e., G_c) to the same degree across the D/HOH and norm groups. However, the Vocabulary subtest had a higher loading on VCI (G_c) for the norm group than for the D/HOH sample. In contrast, the Similarities subtest had a higher loading on VCI (i.e., G_c) for the D/HOH sample than for the norm group. For the D/HOH sample, the VCI latent construct is better predicted by the Similarities subtest than the other two VCI subtests.

The lack of invariance for the VCI subtests across the D/HOH sample and norm group may be a function of measurement error, differences in cognitive ability, differences in cognitive structure or some combination of the three. In considering measurement error, the lack of metric invariance between the norm group and D/HOH sample may be due to the presence of construct-irrelevant variance. Construct-irrelevant variance occurs when the test includes factors extraneous to the construct of interest. These irrelevant factors either serve to increase or decrease the difficulty of the test for a person or a particular group of people. Verbal items are more difficult for D/HOH examinees compared to normal hearing examinees. This is most likely due to the fact that many D/HOH individuals are denied the opportunity to acquire verbal and social knowledge as a direct consequence of their hearing loss (Braden & Hannah, 1998; Maller, 1996). The verbal subtests of the WISC-IV may be measuring different constructs (e.g., lack of verbal experience, differences in item

translations from oral to sign language) for the D/HOH sample than the construct identified for the norm group (i.e., *Gc*). These additional factors may contribute to or be the cause of the lack of invariance found for the VCI subtests between these two groups. “There is uniform agreement that systematic deprivation of exposure to verbal, socially specific knowledge impairs performance on verbal scales independent of an individual’s underlying aptitude” (Braden, 1994, p. 76).

The lack of metric invariance for the VCI subtests may also be due to differences in cognitive ability and/or the way cognitive abilities are structured between D/HOH and hearing children. Environmental influences that result in growing up without hearing may have lasting cognitive effects. Growing up deaf may affect intellectual development, especially language and auditory dependent cognitive processes such as verbal comprehension, verbal reasoning and auditory processing. Therefore, the underlying construct assessed by the VCI subtests may be different for D/HOH children compared to hearing children. Nonstandard exposure to language reduces opportunities for accessing and exploring the surrounding community and culture, which leads to an impoverished knowledge base on which to acquire and build new information (Braden, 2005). The nonstandard exposure to language (at an early, prelingual age) may result in differences in cognitive ability and/or the way cognitive abilities are structured. “Beyond the normal heterogeneity seen in the hearing population, differences in the environments and experiences of deaf children and hearing children might lead to different approaches to learning, to knowledge organized in different ways, and to different levels of skill in various domains. Ignoring this possibility not only denies the reality of growing up deaf in a largely

hearing world, but jeopardizes academic and future vocational opportunities for deaf children” (Marschark, 2003, p. 464). Most research suggests that deaf and hearing individuals vary in their approaches to cognitive tasks and that these differences are likely influenced by the mode of communication. Furthermore, the etiologies of deafness may cause some children to have unidentified cognitive impairments resulting in differences in cognitive ability between D/HOH and hearing children (Maller & Ferron, 1997). For example, meningitis, neurological damage and some syndromic genetic conditions can cause deafness as well as cognitive delays. However, “different” does not necessarily mean “deficient.” I will consider the possibility that differences vs. deficiencies in psychological processes account for my findings in the discussion of hypothesis three.

For D/HOH children lack of verbal experience may lead either to differences in cognitive ability or differences in the structure and organization of cognitive abilities. Furthermore, the nonstandard exposure to language and/or the nonstandard methods of test administration may introduce construct-irrelevant variance. Any or all of these confounding variables may be the reason(s) for the invariant factor loadings of the VCI subtests onto the VCI composite (or second-order CHC factor of Gc) between the two groups. Overall, the VCI subtests do not function in the same way for D/HOH and hearing children, regardless of the underlying cause(s) for these differences. At this time, psychologists should not use the VCI as a measure of cognitive ability for D/HOH children. The number of confounding variables impacting the VCI score, increases the likelihood of misinterpretation and, ultimately, the chances of making ill-informed decisions.

Test of partial metric invariance: PSI. In the third test of partial metric invariance, I constrained only the factor loadings that comprise the PSI to be equal across the two groups (i.e., Coding [CD] and Symbol Search [SS]). Metric invariance was established across the two groups for the PSI subtests. This means that two subtests of the PSI measured the latent factor PSI (i.e., second-order CHC factor G_s) in the same way and to the same degree across the two groups.

The two PSI subtests of the WISC-IV PSI are intended to measure the speed of graphomotor and mental processing (Wechsler, 2003). The Coding and Symbol Search subtests were part of the Performance Scales in the WISC-III; however, they were extracted to form their own Index in the WISC-IV. Although the directions for the two PSI subtests are provided verbally, these two subtests use language reduced items to create, pencil and paper tasks. For the Coding subtest, the examinee uses a key to fill in missing shapes on a grid. For the Symbol Search subtest, the examinee looks through symbols to find matching pairs. For each task, the examinee is timed for two minutes. The Processing Speed (i.e., G_s) CHC factor assesses an individual's ability to perform cognitive tasks fluently and automatically, testing speed and accuracy (Weiss, Prifitera, & Saklofske, 2005). The PSI subtests appear to function similarly to the PRI subtests.

The current finding is somewhat consistent with the research of Maller and Ferron (1997) who investigated the factor structure of the WISC-III for a sample of severely and profoundly deaf children ($N = 110$). Tests of partial invariance revealed that the factor loadings were invariant across the D/HOH sample and the WISC-III norm group ($N = 2200$) for the Symbol Search subtest. However, the factor loadings for the Coding subtest were not

equivalent across groups. The Coding subtest had a higher factor loading for their D/HOH sample (.814) compared to the norm group (.605).

The current finding provides evidence (based on internal structure) for the use of the PSI subtests with D/HOH children. As previously stated, language-reduced measures of intelligence (such as the Coding and Symbol Search subtests) are recommended for assessing the cognitive abilities of D/HOH children (Braden & Athanasiou, 2005; Bradley –Johnson & Evans, 1991; Sullivan & Vernon, 1979; Zieziula, 1982). The current finding that the PSI subtest factor loadings are invariant between the norm group and D/HOH sample suggests that cognitive processes associated with processing speed are similarly structured for D/HOH and hearing children.

Additional Analyses

Due to the fact that partial metric invariance was established for the PRI and PSI subtests, I conducted an additional test of invariance (i.e., scalar invariance). Scalar invariance investigates the equivalence of regression item intercepts (on the latent factors) across groups.

I assessed scalar invariance for the PRI and PSI subtests separately in continuation of tests of partial invariance. The results showed the scales for the PRI or PSI subtests were not invariant (i.e., they were different between the norm and D/HOH groups). Lack of scalar invariance suggests PRI and PSI items function differently for the two groups. Data regarding subtest means were used to test for scalar invariance. The unconstrained intercept values were equivalent to the subtest means. When observing subtest means, the D/HOH sample scaled score subtest means were all lower than those of the norm group, which are

assumed to reflect population values (i.e., $M = 10$; See Table 11.) The lower subtest means could be the result of a systematic bias of test items for D/HOH children or it could mean that D/HOH children (as a group) perform more poorly on the cognitive tasks of the WISC-IV compared to hearing children. Overall, the D/HOH sample scored lower on the PRI and PSI items compared to the norm group. However, the lack of scalar invariance affects the ability to make valid comparison of mean differences between groups.

The lack of scalar invariance found between the D/HOH sample and the norm group may be attributable to error introduced externally during test administration, true differences inherent within the population, or a combination of the two. One possible source of bias affecting all subtests may have been introduced during test administration. It is common practice, outlined in the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 1999) that all individuals are tested in their primary mode of communication. Therefore, a psychologist planning to give the WISC-IV to a D/HOH child who communicates via sign language must administer the assessment in the primary mode of communication - sign language. In the current sample, over 75% of the examinees primarily communicated via sign language (see Table 9) and thus were administered the WISC-IV in that mode of communication. However, simply providing a signed administration of a test is not adequate to guarantee a valid test translation (Braden, 2005). The translation of a test from the source language to a target language requires an extensive, multi-step process. To date there are no fully validated ASL translations (i.e., no normed ASL translations) of any standardized intelligence test, including the WISC-IV (Braden, 2005). The translation of the WISC-IV test directions/items into sign language may have introduced construct-irrelevant

variance or bias that increased the difficulty of the test for this special sample. For some subtests, the translation of directions and/or test items may have altered the target skill (i.e., the skill/latent construct the assessment intends to measure; Braden, 2005), introducing additional variance within test scores.

Furthermore, there is a plethora of evidence to suggest that performance and motor-free nonverbal measures of intelligence do not yield the same results with D/HOH children. Research has shown that D/HOH examinees tend to score lower on nonverbal reasoning tests that reduce or eliminate the role of manual dexterity (e.g., motor-free nonverbal measures) relative to performance tests that require the manual manipulation of items for task completion (Braden, 1994; Braden, 2005; Braden, Kostrubala & Reed, 1994). The PRI subtests of the WISC-IV have undergone the most extensive changes from the WISC-III. The PRI subtests intend to invoke more nonverbal reasoning and less visual spatial skills than its predecessor, the PIQ (Weiss, Prifitera, & Saklofske, 2005). The three new subtests of the PRI are thought to provide a more accurate measure of fluid reasoning and better assess the ability to reason with novel (i.e., less crystallized knowledge) information, while also reducing the emphasis on speed. In fact, even the Block Design subtest (the only performance-like subtest) was modified to decrease the influence of time bonuses (an important characteristic of performance tests). It appears that with the subtest changes, the PRI may no longer function like a performance battery, but instead function more like a motor-free nonverbal battery.

Another reason for the lack of scalar invariance found between the D/HOH sample and the norm group may be the result of unique, specific neuropsychological deficits present

in the current sample. One important neuropsychological deficit to consider is Attention-Deficit Hyperactivity Disorder (ADHD). Thirty-nine examinees in the sample (29.1%) had a co-morbid diagnosis of ADHD. Children diagnosed with ADHD are challenged by tasks measuring selective attention (Schwean & Saklofske, 2005). Prifitera and Dersh (1993) and Schwean, Saklofske, Yackulic and Quinn (1993) examined subtest scores on the WISC-III for children diagnosed with ADHD and found that subtests with the lowest mean scores were those composed of processing speed factors (i.e., Coding, Arithmetic, Symbol Search, Digit Span). Both studies also reported that the lowest mean subtest score for the examinees diagnosed with ADHD was on the Coding subtest. Swanson (2002) also reported that examinees diagnosed with a Learning Disability (LD), ADHD or LD/ADHD scored significantly lower than normally developing children on the Processing Speed factor of the WISC-III. Ten children in the current sample (7.5%) had a comorbid diagnosis of LD in addition to the hearing loss. The Processing Speed tasks on the WISC-IV primarily tap an aspect of cognitive efficiency, reflecting the speed at which a child can make visual symbol discriminations. ADHD children who were administered the WISC-IV obtained their lowest composite score on the PSI, with Coding being the most difficult subtest. Coding also proved to be the most difficult subtest for the sample of ADHD/LD children (Schwean & Saklofske, 2005).

Furthermore, approximately 19% of the sample had a comorbid diagnosis of Spina Bifida Occulta . Although Spina Bifida Occulta is considered the mildest form of Spina Bifida, the malformation of the spine may affect cognitive and motor development/ability. All nerves located below the malformation are affected to some degree. Therefore, the higher

the malformation occurs on the back, the greater the amount of nerve damage and loss of muscle function and sensation (National Institutes of Health, 2011). The extent of the impact of this comorbid health impairment on the study findings is unknown.

Although the subtests of the PRI and PSI appear to load onto their respective latent factors to the same degree across groups, the two groups do not perform in a similar way on these subtests. Lower cognitive abilities in the D/HOH children in my sample may produce the lower mean scores, or may reflect differences in confounding variables (e.g., measurement error, unique sample characteristics) or a combination of both. Because these factors cannot be teased apart, mean differences cannot be solely attributed to differences on the construct of interest.

Hypothesis Three: Differences in Mean Scores

Verbal Comprehension Index (VCI). I predicted that the mean VCI for the D/HOH sample would be lower than the mean VCI reported for the norm group in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003). My hypothesis was supported. The mean VCI for the D/HOH sample ($M = 80.05$) was significantly lower than the population mean ($M = 100$) and this mean difference had a large effect size ($d = 1.31$).

Group comparison studies that test hypotheses about expected differences in average scores across different groups is one way to provide validity evidence for a test (i.e., evidence based on relations with other variables). In fact, validity evidence based on relations to other variables is the most extensive of the five sources of evidence for the validity of a test (Goodwin & Leech, 2003). The current finding that the D/HOH sample had an average VCI score that was significantly lower than the norm group provides evidence for the validity of

the VCI subtests, in that the finding replicates previous research and, therefore, produces expected results. Language-mediated reasoning and knowledge is attenuated by hearing loss. Most D/HOH children (especially deaf children) do not acquire and develop spoken English skills in the same manner as normal hearing children (Maller & Braden, 1993). Therefore, lower scores on a test of verbal knowledge for D/HOH children provide one type of evidence to support the validity of the test. However, the VCI subtests are not recommended for assessing the cognitive abilities of D/HOH children. For this special population, language-loaded intelligence tests confound language skills (or lack thereof) with intelligence. The use of verbal intelligence tests could result in inappropriate consequences for D/HOH individuals (Braden, 2000).

Although some psychologists may believe that the current findings provide evidence to support the validity of the VCI subtests, other psychologists would argue that conducting (and interpreting) group comparisons of mean scores is, in-and-of-itself, invalid when evidence of metric invariance is lacking. Researchers (e.g., Vandenberg & Lance, 2000) argue that, to make valid group comparisons of mean scores, metric (and scalar) invariance should be established. "...It is necessary to show that the two measurements are psychometrically equivalent to make valid group comparisons regarding change over time or across groups of respondents" (Meade & Lautenschlager, 2004, p. 60). Metric invariance was not established for the VCI subtests and, therefore, a test of scalar invariance was unnecessary. The comparison of VCI mean scores between the two groups is likely confounded by construct-irrelevant variance. "Without measurement equivalence, it is difficult to interpret observed mean score differences meaningfully. That is, observed mean

score differences may reflect the true mean difference between the groups as well as a difference in the relationship between the latent variable and the observed score that is not identical across groups (Raju, Laffiette, & Byrne, 2002, p. 517). Although the current findings are consistent with years of previous research, the comparison of means and interpretation of mean differences should be executed with caution. Without evidence of metric (and scalar) invariance, it is not clear as to whether the differences in means are due to real differences, measurement error or some combination of the two. Therefore, mean score comparisons are rendered uninterpretable as representations of a deaf or hard-of-hearing child's intellectual ability.

Despite this argument, verbal scores for D/HOH children may still be clinically useful in a number of different situations. For example, research demonstrates that verbal scores are better predictors of academic achievement and occupational success in D/HOH individuals than nonverbal scores of intelligence (Kelly & Braden, 1990; Maller & Braden, 1993). Also, verbal scores can be useful measures of incidental learning and language acquisition, which can help estimate performance in educational, vocational and social contexts (Braden, 2005). Researchers documented that the Verbal subtests of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) are better at predicting literacy in deaf adults than the Performance subtests (Moores, et al., 1997). It has also been suggested that the WAIS-R Verbal Scale is a better predictor of deaf students' success in college than the Performance Scale (Falberg, 1983). It may be that English language skills underlie both verbal intelligence and academic achievement (Maller & Braden, 1993). Some examiners also use verbal batteries in an attempt to identify D/HOH children with unusual strengths or weaknesses. Verbally gifted

D/HOH children should be exposed to challenging educational experiences and D/HOH children with verbal learning disabilities should receive appropriate support services (Maller, 2003). Although the subtests of the VCI are not recommended to measure the cognitive abilities of D/HOH children, these subtests may be utilized as part of a multi-method evaluation that can guide educational and vocational decisions.

Perceptual Reasoning Index (PRI). Nonverbal measures of intelligence (like the subtests of the PRI) are typically recommended for use with D/HOH individual because they reduce confounds between oral language deficits and intellectual abilities (Braden 1990; Braden Kostrubala & Reed, 1994; Kostrubala, 1998; Maller & Braden, 1993; Murphy, 1957). The mean PRI for the D/HOH sample ($M = 96.18$) was significantly lower than the population mean PRI ($M = 100$). Although this sample of D/HOH children had a mean PRI score that was significantly lower than the population mean, the mean score is still well within the “average” range and the effect size of this difference was small ($d = 0.26$). Therefore, the magnitude of this difference may not be *clinically* significant.

There may be several reasons for the finding that the D/HOH sample had a significantly lower PRI score than the norm group. First, the language demands for the PRI subtests may be greater than previous editions of the WISC. In general, nonverbal tests attempt to reduce language; however, language is not entirely eliminated from the subtests. As a result, verbal abilities may be tapped within these “nonverbal” tasks. For example, the WISC-III subtests (with the exception of Picture Completion) began with the examiner modeling the task. Although verbal directions supplemented the visual display, understanding was likely attained by watching the examiner model the task. In contrast, the

subtests of the PRI do not provide examiner modeling of the task. Two of the three subtests (i.e., Matrix Reasoning and Picture Completion) rely solely on verbal directions to explain the task. The Block Design subtest is the only subtest that begins with a model (supplemented heavily with verbal directions). Therefore, it appears that the PRI subtests rely more heavily on language for task comprehension than previous “nonverbal” portions of the WISC.

Second, the PRI subtests elicit more nonverbal reasoning and reduce visual spatial and motor skills, relative to its predecessor. Compared to the other WISC-IV composites, the PRI has undergone the most extensive changes from the earlier WISC-III PIQ composite (Weiss, Prifitera and Saklofske, 2005). Only one core Performance Scale subtest (i.e., Block Design) from the WISC-III PIQ remains on the WISC-IV PRI. Although the PRI still includes the Block Design subtest, a timed task that requires the manual manipulation of blocks, the other two core PRI subtests (i.e., Picture Concepts and Matrix Reasoning) are *not* timed and do *not* require the manual manipulation of objects (only pointing to correct responses). All of the WISC-III PIQ subtests were either timed (i.e., Picture Completion and Coding) or were timed *and* required the manipulation of objects (i.e., Block Design, Picture Arrangement and Object Assembly). The developers of the WISC-IV claim that the new PRI provides a better measure of fluid reasoning than PIQ because it is less influenced by visualization, manual dexterity and processing speed (Braden, 2005). Previous research has shown that D/HOH examinees tend to score lower on nonverbal reasoning tests that reduce or eliminate the role of manual dexterity relative to performance tests that require the manual manipulation of items for task completion (Braden, 1994; Braden, 2005; Braden, Kostrubala

& Reed, 1994). It appears that with the subtest changes, the PRI may no longer function like a performance battery, but instead function more like a motor-free nonverbal battery. In fact, even the Block Design subtest, although still timed, was modified for the WISC-IV to decrease the influence of time bonuses (an important characteristic of performance tests).

Third, it is possible that the previous PIQs may have simply overestimated g , and that PRI (as a better measure of g) is more accurate. Therefore, the PRI may be a better reflection of the general intelligence of D/HOH children. One goal of the WISC-IV modifications was for the PRI subtests to provide a more pure measure of fluid reasoning that is less influenced by visualization, manual dexterity and processing speed (Braden, 2005). In the WISC-III, five of the seven PIQ subtests (i.e., Block Design, Object Assembly, Picture Arrangement, Picture Completion, and Mazes) loaded heavily onto the CHC second-order “Visual Processing” factor. The other two PIQ subtests (i.e., Coding and Symbol Search) loaded primarily onto the CHC second-order “Processing Speed” factor (Alfonso, Flanagan, & Radwan, 2005). In the WISC-IV, only one of the PRI subtests loads onto the CHC second-order factor “Visual Processing” (i.e., Block Design); the other two subtests load onto “Fluid Intelligence” (i.e., Picture Concepts and Matrix Reasoning). Therefore, it can be assumed that the WISC-IV PRI better represents the second-order “Fluid Intelligence” factor than did PIQ. Within the CHC theory, Fluid Intelligence loads more heavily on g than the Visual Processing and Processing Speed factors. Therefore, one explanation for the lower mean PRI for this sample could be that the new PRI provides a better measure of g than did the WISC-III PIQ. This explanation implies that previous research using WISCs’ PIQ may have overestimated g for D/HOH children.

However, as previously discussed, scalar invariance was not established for the PRI subtests between the D/HOH sample and the norm group (although metric invariance was supported between the two groups). The PRI subtests appear to load onto the PRI second-order factor to the same degree across the D/HOH sample and norm group; however, item intercepts are not equivalent across the two groups making direct comparisons of mean scores questionable. Overall, the PRI is thought to be the best estimate of intelligence, or g , for D/HOH children at this time. However, psychologists should not attribute mean group differences solely to cognitive deficiency until further research is conducted exploring the scalar invariance of the PRI subtests among D/HOH and hearing children.

Processing Speed Index (PSI). I predicted the mean PSI for the D/HOH sample would not differ from the mean PSI reported for the norm group in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003). This hypothesis was not supported. The mean PSI for the D/HOH sample ($M = 94.16$) was significantly lower than the population mean ($M = 100$) however, the effect size of this difference was small ($d = 0.39$).

The PSI is composed of subtests intended to measure the speed of graphomotor and mental processing (Wechsler, 2003). The authors of the WISC-IV claim that these subtests directly measure speed and accuracy, as well as the ability to scan and track simple visual information. Performance on the PSI reveals the rapidity with which an examinee can process simple or routine information without making errors (Weiss, Prifitera, & Saklofske, 2005). The Coding and Symbol Search subtests were part of the Performance Scales in the WISC-III; however, they were extracted to form their own Index in the WISC-IV.

The finding that the D/HOH mean PSI score was significantly lower than the population mean was surprising. Research has found evidence that deaf adults who use sign language show better performance in some aspects of visual perception, such as the ability to rapidly scan visual stimuli or shift visual attention, (Parasnis & Samar, 1985; Rettenback, Dillar, & Sireteanu, 1999) and better detection of motion in the visual periphery, redirection of visual attention from one location to another, and recall of complex visual signs (Corina, Kritchevsky, & Bellugi, 1992; Parasnis & Samar, 1985) than hearing adults and deaf adults who communicate orally. Furthermore, deaf and hearing signers are faster at face recognition (Bellugi et al., 1990), generating visual mental images (Emmorey, Kosslyn, & Bellugi, 1993; Emmorey & Kosslyn, 1996), and visually rotating two-dimensional objects (Emmorey, et al., 1993; Emmorey, Klima, & Hickok, 1998) than hearing peers. However, these advantages are more likely due to the use of visuospatial sign language than the hearing loss per se (Marschark, 2006). Taken together, this research suggests that the possibility that some of the specific stratum I abilities, falling under the stratum II *Gs* factor (e.g., perceptual speed and rate of test taking) may be better developed in deaf adults and children than in their hearing peers.

However, other research shows that processing speed is related to reading performance and development (Kail & Hall, 1994), the efficient use of working memory for higher order fluid tasks (Fry & Hale, 1996; Kail 2000) and overall mental capacity (Kail & Salthouse, 1994). Differences between the D/HOH sample and the norm group on one or more of these variables may attribute to the PSI mean differences found in this study.

Similarly, Braden (1990) conducted a meta-analysis of 21 studies and found that on Wechsler

Performance Scales, deaf individuals consistently scored lower on the Coding/Digit Symbol subtests relative to other Performance subtests. Braden suggested that low Coding/Digit Symbol scores could be the result of neurological impairments more common in the deaf population. Following this logic, the PSI score may be attenuated by subtle neuropsychological deficits present in this D/HOH sample (see discussion of ADHD and Spina Bifida Occulta above).

Although this sample of D/HOH children had a mean PSI score that was significantly lower than the population mean, the mean score ($M = 94.16$) is still well within the “average” range. Therefore, despite the presence of statistically significant findings, the WISC-IV PSI may still provide clinically significant information that can help psychologists, educators and parents make informed decisions about educational accommodations, interventions and placements. Lower scores on Processing Speed subtests have been associated with learning problems in D/HOH children, but correlation cannot imply causation (Braden, 2005). “Lower scores may indicate less-developed cognitive processes, untreated or related conditions (e.g., attention deficit disorder), or lack of learning opportunities (e.g., delayed exposure to writing, the alphabet, and numerals may influence performance on tests independent of cognitive abilities” (Braden, 2005, p. 371).

However, as previously discussed, scalar invariance was not established for the PSI subtests between the D/HOH sample and the norm group (although metric invariance was supported between the two groups). The PSI subtests load onto the PSI second-order factor to the same degree across the D/HOH and norm groups; however, item intercepts are not equivalent across the two groups. Therefore, psychologists should not make direct

comparisons of mean scores between groups as mean group differences may not reflect true mean differences on the construct of interest. However, the PSI may be helpful in identifying relative strengths and weaknesses (using ipsative comparisons of scores) in cognitive processing that may affect educational and vocational performance.

Hypothesis Four: Subtest Intercorrelations

I hypothesized that the correlations among the WISC-IV subtests would be positive and reliably greater than zero. My prediction was based on the assumption that all the subtests of the WISC-IV should measure general intelligence (i.e., *g*) so, therefore, it was expected that they would positively correlate with one another. Of the 28 correlations between the available WISC-IV subtests, 3 correlation confidence intervals included zero. Investigating these non-significant correlations, all three correlations included the Vocabulary subtest of the VCI and one other non-VCI subtest (i.e., Block Design, Symbol Search and Coding). Although the VCI subtests correlated highly with each other, on average, they had only a weak correlation with the other subtests. The Vocabulary subtest was the only VCI subtest with nonsignificant correlations. All subtests within the same index were significantly correlated and all non-significant correlations were between-index correlations.

When considering the standardized regression weights for the D/HOH and norm groups (see Figures 4 and 5), the Vocabulary subtest had the highest loading on VCI for the norm group (.89) and the lowest loading on VCI for the D/HOH sample (.74) compared to the other two subtests (Similarities and Comprehension). When considering the VCI subtests,

the Vocabulary functions most differently between the two groups which likely contributed to the nonsignificant correlations.

The Vocabulary subtest is designed to measure a child's word knowledge and verbal concept formation. However, it also measures a child's fund of knowledge, learning ability, long-term memory and degree of language development (Wechsler, 2003). Compared to the other two subtests of the VCI (i.e., Similarities and Comprehension), the Vocabulary subtest evokes the least amount of reasoning (Weiss, Prifitera, & Saklofske, 2005). "The Vocabulary (VC) subtest requires that the meaning of a word was learned, can be recalled, and expressed coherently. There is no apparent demand to reason in this subtest" (Weiss, Prifitera, & Saklofske, 2005, p. 73). For the norm group (thought to reflect the hearing population), the Vocabulary subtest is one of the highest "g" loaded subtests and one of the best predictors of overall intelligence. High-order thinking requires chunking large amounts of information into a coherent whole. Individuals with larger vocabularies can chunk more concepts into a single word than individuals with smaller vocabularies (Weiss, Prifitera & Saklofske, 2005). A diverse vocabulary is directly related to the exposure to the English language. Children with larger vocabularies likely enjoyed a more enriched cognitive environment than children with small, less diverse vocabularies.

Children with deficits in crystallized knowledge may score higher on the Similarities and Comprehension subtests than the Vocabulary subtest if they have adequate verbal reasoning ability (Weiss, Prifitera & Saklofske, 2005). The mean score for the Vocabulary subtest for the D/HOH sample ($M = 4.75$, $SD = 3.10$) was lower than the mean scores for the Similarities ($M = 7.54$, $SD = 3.24$) and Comprehension ($M = 7.63$, $SD = 4.44$) subtests. The

Similarities and Comprehension subtests also had positive, significant correlations with all the other subtests assessed, suggesting they are all measuring a similar construct (i.e., *g*). However, the average strength of the correlation coefficients among one of the VCI subtests and a non-VCI subtests was low (.25), compared to the average correlation coefficient (.37) for non-VCI subtests (i.e., Block Design, Matrix Reasoning, Picture Completion, Symbol Search and Coding). These differences suggest that the Vocabulary subtest may not be measuring verbal reasoning ability for the D/HOH sample, but instead be a more direct measure of language (e.g., vocabulary) deficits for this special population than the other two subtests of the VCI.

Summary

The results of the present study show that the CHC theory, as realized through the WISC-IV, is an appropriate theory of cognitive structure for D/HOH children. Overall, the data for the D/HOH sample fit the factor structure of the WISC-IV, suggesting the presence of the same number of indexes (i.e., PRI, VCI and PSI)/ CHC second-order factors (i.e., *Gc*, *Gf/Gv*, *Gs*) for D/HOH and hearing children. However, there are important differences within the WISC-IV composites that should be considered for both theoretical and practical reasons.

First, the subtests measuring VCI (or *Gc*) do not function in the same way and to the same degree for D/HOH and hearing children. For hearing children, the VCI subtests assess culturally specific knowledge and verbal reasoning skills. For D/HOH children, the VCI subtests are confounded with the hearing loss and, instead of assessing culturally specific knowledge and verbal reasoning skills, assess the lack of exposure to oral language. For D/HOH children, the subtests that are intended to measure *Gc* (i.e., Similarities, Vocabulary

and Comprehension) confound intelligence with language/cultural knowledge deficits resulting from the hearing loss. The Vocabulary subtest relies more on culturally acquired knowledge and less on reasoning skills than the other two subtests (i.e., Similarities and Comprehension). Furthermore, D/HOH children score significantly lower on the VCI composite than hearing children. However, this mean comparison is called into question due to the lack of metric invariance found for the VCI subtests. Overall, psychologists should not use the VCI subtests as a measure of cognitive ability for D/HOH children. However, the VCI may provide valuable information in some situations as a measure of incidental learning and language acquisition.

Second, the PRI and PSI subtests load onto their respective constructs (i.e., Gf/Gv and Gs) in the same way and to the same degree for D/HOH and hearing children. However, I found a lack of scalar invariance within both composites (i.e., PRI and PSI). Therefore, the item intercepts on these variables are not the same between D/HOH and hearing children. The items of the PRI and PSI are systematically downward biased for the sample of D/HOH children. This downward bias may be attributable to several reasons such as differences in cognitive ability, presence of measurement error, unique characteristics of the sample or a combination of all three. Without evidence of scalar invariance, a direct comparison of means scores between different groups of children is not supported. The D/HOH sample scored significantly lower on the PRI and PSI composites compared to the norm group (thought to reflect population values). However, the mean scores remained within the average range. Therefore, these differences may not be clinically significant. Overall, the

second-order Gf/Gv and Gs abilities appear to be organized in the same way for D/HOH and hearing children and can be used for measuring g .

Implications for Practice

In addition to helping advance understanding of the organization of cognitive abilities within a CHC framework, my study also informs assessment practice with D/HOH children. Standard 11.16 states, “Test users should verify periodically that their interpretations of test data continue to be appropriate, given significant changes in their population of test takers, their modes of test administration, and their purposes in testing” (AERA, APA, & NCEM, 1999, p. 117). Therefore, it is important to understand how the current findings can be used to guide the practice of psychologists, educators and others who work with D/HOH children.

The PRI and PSI subtests load onto the WISC-IV Indexes PRI and PSI, respectively, in the same way and to the same degree for both D/HOH and hearing children. In other words, these subtests appear to measure the second-order CHC factors Gf/Gv and Gs in the same way for D/HOH and hearing children. This finding is helpful because language-reduced tests are widely used with this special population. This finding also provides evidence (based on internal structure) for the validity of these tests for D/HOH children.

However, lack of scalar invariance for the PRI and PSI subtests suggests the presence of test or item bias for these scales. The D/HOH subtest means were all lower than the subtest scaled score mean of 10. Furthermore, the PRI and PSI means are both significantly lower than the standard score mean of 100, which is thought to reflect population values. Subtest items appear to be more difficult for D/HOH children than hearing children, resulting in a systematic “downward bias” for this special population.

Practitioners should understand that low PRI and PSI scores may not solely reflect a D/HOH child's intelligence, but are likely confounded by the types of tasks given, the language required to understand task directions, and the alternate mode of implementation (e.g., ASL). Low scores may also be due to specific neuropsychological deficits such as ADHD, Spina Bifida Occulta, a learning disability or any number of additional comorbid diagnoses. These confounding variables alone or in interaction with each other may cause spuriously low scores that attenuate the construct of interest (i.e., *Gf/Gv*, *Gs*, and ultimately, *g*) for this population. In general, language reduced measures of intelligence are better to use when estimating the intelligence of D/HOH children than language-loaded measures. The recommendation to use language reduced measures is consistent with previous research conducted on the Wechsler Scales and supports the current assessment practices of most psychologists working with this special population.

The three subtests of the VCI measure the CHC second-order factor *Gc* in a different way and to a different degree with D/HOH children as compared to hearing children. This finding supports the general consensus that verbal tests of intelligence should not be used to assess the cognitive abilities of D/HOH children. The WISC-IV VCI should not be used as a measure of intelligence for D/HOH children. The WISC-IV VCI confounds language skills with intelligence and, therefore, could result in inappropriate consequences for D/HOH children if used as an assessment of cognitive ability.

In general, practitioners should not interpret low scores as evidence of deficits in cognitive processes until they can reasonably rule out other competing explanations. Cognitive interpretations of the WISC-IV test scores should be avoided with D/HOH

children because differences in language and auditory experiences may be more plausible explanation of low scores than neuropsychological processes. Examiners should consult with experts familiar with the intellectual assessment of D/HOH children before drawing solid conclusions regarding unusual WISC-IV score profiles. As is true for any case, the WISC-IV should only be one source of data compiled from a multi-method, multi-informant evaluation. There is no evidence to justify specific educational or psychological interventions on the basis of WISC-IV scores alone (Braden, 2005).

The valid, proper use of tests can help educators and psychologists make informed decisions about children that can lead to more equitable access to education. However, improper use of tests can cause serious detrimental consequences to test takers and other parties affected by test-based decisions (AERA, APA, & NCME, 1999). When conducting an assessment, psychologists are ethically obligated to choose the best available procedures (Jacob & Hartshorne, 2003). Psychologists should also understand that many of our current testing procedures rely on assumptions of measurement equivalence/invariance across different populations. “Under many conditions of research, evidence of invariance of measurement is necessary for drawing clear inference from results. Such clear inferences must be at the foundation of valid scientific explanation” (Horn & McArdle, 1992, p. 118). Based on the findings of the current study, mean score differences may reflect true ability differences, may be a product of confounding variables related to measurement error or may reflect a combination of these two possibilities. Therefore, psychologists must understand that the interpretation of the WISC-IV test scores requires clinical judgment that must be exercised on a case-by-case basis using all of the data available.

Limitations

There are several limitations to this study. The most obvious limitation is the fact that the present study excluded the WMI subtests from the factor analyses. The significant challenges in collecting this information inhibited me from conducting a full factor analysis of the WISC-IV. The current findings are therefore limited with respect to guiding present practice.

A second notable limitation is sample size. The present study did not meet the target sample size of 150 ($N = 134$). Some researchers (e.g., Guilford, 1954; Cattell, 1978; Comrey & Lee, 1992; MacCallum, Widaman, Zhang & Hong, 1999) would consider this sample size too small for conducting tests of factor analysis. In general, larger samples produce more reliable results; therefore, some may question the reliability of the current findings. However, other researchers believed the sample size to variable ($N:p$) ratio was of most importance. Cattell (1978) believed this ratio should be 3-6:1. Gorsuch (1983) argued for a minimum ratio of 5:1 and Everitt (1975) recommended a ratio of at least 10:1 (MacCallum, et al., 1999). Field (2005) said the most common ‘rule of thumb’ is to have 10-15 participants per variable. MacCallum et al. (1999) believed that “the minimum level of N , or the minimum $N:p$ ratio, needed to assure good recovery of population factors is not constant across studies, but rather is dependent on some aspects of the variables and design in a given study” (p. 96). Considering the most stringent recommendation for sample size made above (i.e., 15 participants per variable), I would need a sample size of 120 (i.e., 15 x 8 variables = 120). As my sample size exceed this recommendation by 14 participants, my sample size should be considered sufficient for the factor analyses conducted in this study.

However, it is also important to understand how the limited sample size impacts the significance testing of the subtlest intercorrelations. Statistically, smaller sample sizes produce larger confidence intervals, which are more likely to include zero. Therefore, the limited sample size may have increased the odds of getting nonsignificant correlations. Due to the fact that I found only three nonsignificant correlations, this does not appear to be a true limitation. Conversely, the large norm group sample size greatly increases the power of the chi-square test to detect small differences. This extreme sensitivity to detect small effects increases the chances of reporting “false positive” results. In other words, the large sample size increases the chances of reporting significantly different findings when, practically, these differences are not meaningful.

A third limitation of this study is the fact that the sample is a “convenience sample” and cannot be considered random. Only when a sample is random can it be considered representative of the population of interest. Therefore, the D/HOH sample used in this study may not be representative of D/HOH children as a group. There may be some characteristic of this sample (related to recruitment procedures and/or availability of data) that makes it different from the population of D/HOH children. For example, almost 30% of the sample had a comorbid diagnosis of ADHD and approximately 19% of the sample had a comorbid diagnosis of Spina Bifida Occulta (as discussed above). The percentage of children diagnosed with ADHD in the general population is approximately 8% (Center for Disease Control and Prevention, 2012) compared to 5.4% among D/HOH children (Gallaudet Research Institute, 2011). The additional neuropsychological factors in the current sample may have skewed study findings (e.g., lower mean PSI scores) as discussed previously. In

addition, the majority of the children in this sample (over 75%) were classified as severely or profoundly deaf (i.e., hearing loss > 70 dB) compared to 39.4% in the general D/HOH population (Gallaudet Research Institute, 2011). The current sample also had an overrepresentation of Hispanic children (40.3%) compared to 25.3% reported for the national data of D/HOH children (Gallaudet Research Institute, 2011). Conversely, the sample underrepresented Black children (5.2%) compared to 14.8% reported for the national data of D/HOH children (Gallaudet Research Institute, 2011). Researchers and practitioners should, therefore, exercise caution when extrapolating these findings to the general population and to children with mild or moderate degrees of hearing loss.

A fourth limitation to this study deals with the data collection process. I had no control over the testing environment and other administration conditions. I assumed that examiners followed all WISC-IV standardization procedures for test administration and scoring, but cannot guarantee these conditions were met. Characteristics of the testing environment (e.g., noise level, lighting, visual distractions) as well as examinee characteristics during the assessment (e.g., fatigue, difficulty attending to tasks, motivation) were not reported nor analyzed. In addition, a fully standardized ASL translation for WISC-IV test directions and items do not exist. Therefore, differences in test administration among the nine study participants are likely. The possible influence of these characteristics on the test scores and interpretations are unknown. Also, because psychologists entered the WISC-IV data into spreadsheets, there was no way to check for entry errors against the original protocols. It is possible that there were entry errors that went unnoticed by the participants and influenced study results.

Finally, findings for this study are largely based on goodness-of-fit statistics. However, there are many model fit statistics described in the literature with multiple different “recommendations” for cutoff values. Therefore, I may have garnered different findings had I used alternate goodness-of-fit statistics or more conservative cutoff values.

Directions for Future Research

Standard 9.2 states, “When credible research evidence reports that test scores differ in meaning across subgroups of linguistically diverse test takers, then to the extent feasible, test developers should collect data for each linguistic subgroup studied in the same form of validity evidence collected for the examinee population as a whole” (AERA, APA, & NCEM, 1999, p. 97). Future research is needed to continue to investigate the factor structure of the WISC-IV with D/HOH children. Conducting a similar study with a larger sample size will increase the confidence in study findings. Guilford (1954) argued that N should be at least 200 and Cattell (1978) said the minimum N should be 250. Comrey and Lee (1992) urged researches to have a sample size of at least 500 and offered the following qualitative scale for factor analysis sample sizes: 100 = poor, 200 = fair, 300 = good, 500 = very good, 1,000 or more = excellent (MacCallum, Widaman, Zhang, & Hong, 1999).

The current study utilized a convenience sample, which may not be representative of the D/HOH population. Therefore, future research is needed to conduct a similar study using a random sample of D/HOH children. Using a random sample of D/HOH children will increase the external validity of the study and allow for more confident generalizations of study findings to D/HOH children as a group.

Future research is needed to further investigate differences in the PRI and PSI subtests for D/HOH and hearing children. The D/HOH sample performed significantly lower on the PRI and PSI than the norm group. In addition, I found lack of scalar invariance across groups for these two composites. More research is needed to determine if differences reported in this study are replicable with more representative, larger samples of D/HOH children.

Future research is also needed to investigate the factor structure of the working memory subtests of the WISC-IV for D/HOH children. Due to significant challenges collecting the data, I was unable to conduct a four-factor analysis of the WISC-IV. Although there are several implications for administering the Working Memory subtests of the WISC-IV (Digit Span and Letter-Number Sequencing), it is likely that some psychologists are using (and interpreting) these subtests as part of psychoeducational evaluations and, therefore, understanding how these subtests function for this population is important. In addition, there are several research findings regarding the visual-spatial and working memory abilities of D/HOH individual that make this aspect of cognitive functioning important to study (see Alfonso, Flanagan & Radwan, 2005; Marschark, 1996; 2003; 2006 ; Marschark & Mayer, 1998).

Future research is needed to help illuminate the “deficiency vs. discrepancy” issues that plague the interpretation of mean score differences. In other words, are the differences found between the D/HOH sample and norm group due to true cognitive differences, differences in measurement error or both? This research is needed to help provide more

practical advice to psychologists who use the WISC-IV as part of a psychoeducational evaluation of D/HOH children.

Conclusions

The primary goal of this study was to understand the intersection of the CHC theory and how it applies to D/HOH children as viewed through the lens of the WISC-IV. Overall, the cognitive structure of the CHC theory is an appropriate model to use with D/HOH children. The cognitive abilities of D/HOH children are organized in the same way compared to WISC-IV norm group. However, for D/HOH children, the subtests of the VCI do not measure the construct of *Gc* to the same degree as hearing children. For D/HOH children, the subtests of the VCI are confounded by language deficits that are a direct consequence of their hearing loss. In addition, the D/HOH sample scored significantly lower on the VCI compared to the norm group (thought to reflect the hearing population). In general, the VCI subtests are not recommended for the cognitive assessment of D/HOH children (a conclusion that is strongly supported by years of research).

The nonverbal subtests of the WISC-IV (i.e., the subtests of the PRI and PSI) are recommended for the cognitive assessment of D/HOH children (another conclusion that is supported by years of research). The subtests of the PRI and PSI measure the constructs of *Gf/Gv* and *Gs* in the same way and to the same degree for D/HOH and hearing children. However, I found a lack of scalar invariance for PRI and PSI, calling into question the ability to make valid comparisons of group means. Although this finding is statistically significant, it may not be clinically relevant. The mean scores for the PRI and PSI for the D/HOH sample were significantly lower than the norm group. However, the mean scores were still well

within the average range. Due to a number of possible confounding variables in this study, it is unsure if the differences in scalar invariance and mean scores are a product of true differences among D/HOH and hearing children or an artifact of measurement error, unique sample characteristics or a combination of all three.

Despite the presence of these significant statistical differences, the WISC-IV is still thought to be a clinically relevant tool to use in the evaluation of and to aid in the educational planning for D/HOH children. For example, the PSI composite may be useful in identifying deficits in cognitive processes that may influence learning and behavior. The VCI may be best understood as an indicator of incidental learning, and clinically, in most cases, the PRI continues to be the best estimate of cognitive functioning for D/HOH children.

References

- Alfonso, V.C., Flanagan, D. P., & Radwan S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues, second edition* (pp. 185- 202). New York: Guilford Press.
- Alwin, D. F., & Jackson, D. J. (1981). Applications of simultaneous factor analysis to issues of factorial invariance. In D. J. Jackson & E.F. Borgatta (Eds.), *Factor Analysis and measurement in sociological research* (pp. 249-279). Beverly Hills, CA: Sage.
- Arbuckle, J., and Wothke, W. (1999). *AMOS 4.0 user's guide*. Chicago: Smallwaters.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*, (pp. 47-89). New York: Academic Press.
- Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, 1, 45-87.
- Banks, J., Gray, C., & Fyfe, R. (1990). The written recall of printed stories by severely deaf children. *British Journal of Educational Psychology*, 60, 192-206.
- Bellugi, U., O'Grady, L., Lillo-Martin, D., O'Grady, M., van Hoek, K., & Corina, D. (1990). Enhancement of spatial cognition in deaf children. In V. Volterra and C. J. Erting (Eds.), *From gesture to language in hearing and deaf children* (pp. 278-298). New York: Springer-Verlag.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.

- Bickley, P. G., Keith, T. Z., & Wolfe, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence, 20*, 309-328.
- Bienenstock, M. A., & Vernon, M. (1994). Classification by the states of deaf and hard of hearing students: It's not how much you hear but where you live. *American Annals of the Deaf, 139*, 128-131.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.
- Braden, J. P. (1984). The factorial similarity of the WISC-R performance scale in deaf and hearing samples. *Journal of Personality and Individual Differences, 5*, 405-409.
- Braden, J. P. (1989). The criterion-related validity of the WISC-R performance scale and other nonverbal IQ tests for deaf children. *American Annals of the Deaf, 134*, 329-332.
- Braden, J. P. (1990). Do deaf persons have a characteristic psychometric profile on the Wechsler Performance Scales? *Journal of Psychoeducational Assessment, 8*, 518-526.
- Braden, J.P. (1994). *Deafness, Deprivation, and IQ*. New York: Plenum Press.
- Braden, J. P. (2000). Deafness and hearing loss. In A. E. Kazdin (Ed.) *Encyclopedia of psychology* (vol. 2) (pp. 441-444). Washington, DC/Oxford: American Psychological Association/Oxford University Press.
- Braden, J.P. (2005). Hard-of-hearing and deaf clients: Using the WISC-IV with clients who are hard-of-hearing or deaf. In A. Prifitera, D.H. Saklofske, L.G. Weiss (Eds.), *WISC-*

- IV clinical use and interpretation* (pp. 351-376). Burlington, MA: Elsevier Academic Press.
- Braden, J. P., & Hannah, J. M. (1998). Assessment of hearing impaired and deaf children with the WISC-III. In D. Saklofske & A. Prifiter (Eds.), *Use of the WISC-III in clinical practice* (pp. 175-201). Boston, MA: Houghton Mifflin.
- Braden, J. P., Kostrubala, C.E., & Reed, J. (1994). Why do deaf children score differently on performance vs. motor-reduced nonverbal intelligence tests? *Journal of Psychoeducational Assessment*, 12, 357-363.
- Braden, J. P., & Shaw, S. R. (2009). Intervention validity of cognitive assessment: Knowns, unknowables and unknowns. *Assessment for Effective Intervention*, 34, 106-115.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications, Inc.
- Byrne, B. M. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. New York: Springer Verlag.
- Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic concepts, applications and programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: basic concepts, applications, and programming*. New York: Routledge.
- Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.

- Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor analytic studies*. New York: Cambridge University Press.
- Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5-22), Amsterdam: Pergamon Press.
- Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues, second edition* (pp. 69-76). New York: Guilford Press.
- Cattell, R. B. (1941). *Some theoretical issues in adult intelligence testing*. Psychological Bulletin, 38, p. 592.
- Cattell, R. B. (1963). *Theory of fluid and crystallized intelligence: A critical experiment*. Journal of Educational Psychology, 54, 1-22.
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Cattell, R. B. (1978). *The scientific use of factor analysis*. New York: Plenum.
- Center for Disease Control and Prevention. (2012). Retrieved March 10, 2012 from <http://www.cdc.gov/nchs/fastats/adhd.htm>
- Cole, D. A., & Maxwell, S. E. (1985). Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach. *Multivariate Behavioral Research*, 20, 389-417.
- Comrey, A. L., Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

- Corina, D. P., Kritchevsky, M., & Bellugi, U. (1992). Linguistic permeability of unilateral neglect: Evidence from American Sign Language. In *Proceedings of the Cognitive Science Conference* (pp. 384-389). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Council for Exceptional Children. (2005). Retrieved October 15, 2005, from <http://www.cec.sped.org/>
- Emmorey, K., Klima, E. S., & Hickok, G. (1998). Mental rotation within linguistic and nonlinguistic domains in users of American Sign Language. *Cognition*, *68*, 221-226.
- Emmorey, K., & Kosslyn, S. (1996). Enhanced image generation abilities in deaf signers: A right hemisphere effect. *Brain and Cognition*, *32*, 28-44.
- Emmorey, K., Kosslyn, S., & Bellugi, U. (1993). Visual imagery and visual-spatial language: Enhanced imagery abilities in deaf and hearing ASL signers. *Cognition*, *46*, 139-181.
- Evans, L. (1980). WISC Performance Scale and Coloured Progressive Matrices with deaf children. *Journal of Educational Psychology*, *50*, 216-222.
- Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry*, *126*, 237-240.
- Falberg, R. M. (1983). Psychological assessment of the verbal functioning of postsecondary educational program applicants: Improving predictive validity. In D. Watson, G. Anderson, P. Marut, S. Ouellette, & N. Ford (Eds.), *Vocational evaluation of hearing-impaired persons: Research and practice*. Warm Springs, AR: University of Arkansas.

- Flanagan, D. P., McGrew, K. S., & Ortiz, S. (2000). *The Wechsler Intelligence Scales of Gf-Gc Theory: A contemporary approach to interpretation*. Needham Heights, MA: Allyn & Bacon.
- Fry, A.F., & Hale, S. (1996). Processing speed, working memory and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*(4), 237-241.
- Furth, H. (1966). *Thinking without language: Psychological implications of deafness*. New York: Free Press.
- Galton, F. (1865). Hereditary talent and character. *Macmillan's Magazine, 12*, 157-166, 318-327.
- Gallaudet Research Institute (2011). 2009-2010 regional and national summary report of data from the 2009-10 annual survey of deaf and hard of hearing children and youth. Washington, DC: GRI. Available from http://www.gallaudet.edu/Gallaudet_Research_Institute/Demographics.html.
- Gerbing, D. W., & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika, 52*, 99-111.
- Goodwin, L.D., & Leech, N.L. (2003). The meaning of validity in the new Standards for Educational and Psychological Testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development, 36*, 181-191.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Guilford, J. P. (1965). *Fundamental statistics in psychology and education*. New York: McGraw Hill.

- Hitch, G. J. (1985). Short-term memory and information processing in humans and animals: Towards an integrative framework. In L.G. Nilsson & T. Archer (Eds.), *Perspectives on learning and memory* (pp. 119-136). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement, 65*, 202-226.
- Horn, J. L. (1965). Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities. Unpublished doctoral dissertation, University of Illinois, Champaign.
- Horn, J. L. (1968). *Organization of abilities and the development of intelligence*. *Psychological Review, 75*, 242-259.
- Horn, J. L. (1976). *Human abilities: A review of research and theory in the early 1970s*. *Annual Review of Psychology, 27*, 437-485.
- Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. J. Woodcock, *WJ-R technical manual* (pp. 197-232). Chicago: Riverside.
- Horn, J. L., & Cattell, R. B. (1966). *Refinement and test of the theory of fluid and crystallized general intelligences*. *Journal of Educational Psychology, 57*(5), 253-270.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117-144.

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

Hu, L.T., and Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structures with missing data using complete data routines. *Structural Equation Modeling*, 6, 1-55.

Individuals with Disabilities Education Act (2006). Retrieved November 10, 2008 from

https://www.ideadata.org/arc_toc8.asp#POP

Joreskog, K.G., and Sorbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.

Jenson, A.R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger Publishers.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.

Kail, R. (2000). Speed of information processing: Developmental change and links to intelligence. *Journal of School Psychology*, 38(1), 51-61.

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology*, 30(6), 949-954.

Kail, R., & Salthouse, T.A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, 86, 199 – 225.

Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2005). A history of intelligence test interpretation. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 23-38). New York: Guilford Press.

- Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley.
- Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. . In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 581-614). New York: Guilford Press.
- Kline, P. (1979). *Psychometrics and psychology*. London: Academic Press.
- Kostrubala, C. (1998). Translation of the WAIS-III from English to ASL. Unpublished master's thesis. University of Wisconsin, Madison.
- Krouse, H. (2008). *The reliability and validity of the WISC-IV with D/HOH children*. Unpublished master's thesis, North Carolina State University, Raleigh.
- Long, D. A., & Perkins, D. D. (2003). Confirmatory factor analysis of the sense of community index and development of a brief SCI. *Journal of Community Psychology*, 31, 279-296.
- Long, D. A., & Perkins, D.D. (2007). Community social and place predictors of sense of community: A multilevel and longitudinal analysis. *Journal of Community Psychology*, 35, 563-581.
- MacCallum, R.C., Browne, M.W., and Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130-149.
- MacCallum, R. C., Widaman, K. R., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.

- Mackintosh, N. J. (1998). *IQ and human intelligence*. New York: Oxford University Press.
- MacSweeney, M., Campbell, R., & Donlan, C. (1996). Varieties of short-term memory coding in deaf teenagers. *Journal of Deaf Studies and Deaf Education, 1*, 249-262.
- Maller, S. J. (1996). WISC-III verbal item invariance across samples of deaf and hearing children of similar measured ability. *Journal of Psychoeducational Assessment, 14*, 152-165.
- Maller, S. J. (1997). Deafness and WISC-III item difficulty: Invariance and fit. *Journal of School Psychology, 35*, 299-314.
- Maller, S. J. (2003). Intellectual assessment of deaf people. In M. Marschark & P.E. Spencer (Eds.), *Oxford handbook of deaf studies, language and education* (pp. 451-463). New York, NY: Oxford University Press.
- Maller, S.J., & Braden, J.P. (1993). The construct and criterion-related validity of the WISC-III with deaf adolescents. *Journal of Psychoeducational Assessment, Advances in Psychoeducational Assessment, Wechsler Intelligence Scale for Children, Third Edition [Monograph], 11*, 105-113.
- Maller, S. J., & Ferron, J. (1997). WISC-III factor invariance across deaf and standardization samples. *Educational and Psychological Measurement, 57*, 987-994.
- Marschark, M. (2003). Cognitive functioning in deaf adults and children. In M. Marschark & P.E. Spencer (Eds.), *Oxford handbook of deaf studies, language and education* (pp. 464-477). New York, NY: Oxford University Press.
- Marschark, M. (2006). Intellectual functioning of deaf adults and children: Answers and questions. *European Journal of Cognitive Psychology, 18*, 70-89.

- Marschark, M., DeBeni, R., Polazzo, M. G., & Masteller, A. (1993). Deaf and hearing-impaired adolescents' memory for concrete and abstract prose: Effects of relational and distinctive information. *American Annals of the Deaf, 138*, 31-39.
- Marschark, M., & Everhart, V.S. (1999). Problem solving by deaf and hearing children: Twenty questions. *Deafness and Education International, 1*, 63-79.
- Marschark, M., & Mayer, T. (1998). Mental representation and memory in deaf adults and children. In M. Marschark & M. D. Clark (Eds.), *Psychological perspectives on deafness, Vol. 2* (pp. 53-77). Mahwah, NJ: Lawrence Erlbaum Associates.
- Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*, 181-220.
- Mayberry, R. I., & Eichen, E. B. (1991). The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. *Journal of Memory and Language, 30*, 486-512.
- McCallum, R. S. (2003). Context for nonverbal assessment of intelligence and related abilities. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 3-22). New York: Kluwer Academic/Plenum.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues, second edition* (pp. 136-181). New York: Guilford Press.
- McGrew, K. S., & Woodcock, R. W. (2001). Woodcock-Johnson III technical manual. Itasca, IL: Riverside.

- Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling, 14*, 611-635.
- Meade, A. W., Johnson, E. C., & Braddy, P.W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*, 568-592.
- Meade, A. W., & Lautenschlager, G. J. (2004a). A monte-carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling, 11*, 60-72.
- Meade, A.W., & Lautenschlager, G. J. (2004b). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7*, 361-388.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.
- Millsap, R.E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology, 73*, 574-584.
- Moores, D., Kluwin, T., Johnson, R., Cox, P., Blennerhassett, L., Kelly, L., Sweet, C., & Fields, L. (1997). *Factors predictive of literacy in deaf adolescents of deaf parents: Factors predictive of literacy in deaf adolescents in total communication programs*

- (Contract No. NO1-NS-4-2365). Washington: National Institute of Neurological and Communicative Disorders and Stroke.
- Mulaik, S. A. (1975). Confirmatory factor analysis. In D. J. Amick & H. J. Walberg (Eds.), *Introductory multivariate analysis* (pp. 170-207). Berkeley, CA: McCutchan.
- Murphy, L. J. (1957). Tests of abilities and attainments. In Ewing, A. W. G. (Eds.), *Educational Guidance and the Deaf Child*. Manchester University Press.
- Mykelbust, M. (1960). *The psychology of deafness*. New York: Grune and Stratton.
- National Center on Birth Defects and Developmental Disabilities (2004). *Hearing loss*. Retrieved October 15, 2005 from <http://www.cdc.gov/ncbddd/dd/ddhi.htm>
- National Dissemination Center for Children with Disabilities (2004). *Deafness & hearing loss: Disability fact sheet*. Retrieved April 8, 2006, from <http://www.nichcy.org/pubs/factshe/fs3.pdf>
- National Institutes of Health (2011). *National Institute of Neurological Disorders and Stroke Spina Bifida Fact Sheet*. Retrieved January 14, 2012 from http://www.ninds.nih.gov/disorders/spina_bifida/detail_spina_bifida.htm
- Neville, H. J., & Lawson, D. (1987). Attention to central and peripheral visual space in a movement detection task: an event-related potential and behavioral study: III. Separate effects of auditory deprivation and acquisition of a visual language. *Brain Research*, 405, 284-294.
- Northern, J. L., & Downs, M. P. (2002). *Hearing in children* (5th ed). Baltimore, MD: Lippincott Williams & Wilkins.

- Ortiz, S. O., & Dynda, A. M. (2005). Use of intelligence tests with culturally and linguistically diverse populations. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues, second edition* (pp. 545-556). New York: Guilford Press.
- Ottum, E. (1980). An analysis of cognitive studies with deaf subjects. *American Annals of the Deaf, 125*, 564-575.
- Parasnis, I., & Samar, V. J. (1985). Parafoveal attention in congenitally deaf and hearing young adults. *Brain and Cognition, 4*, 313-327.
- Pintner, R., Einsenson, J., & Stanton, M. (1941). *The psychology of the physically handicapped*. New York: Crofts.
- Pintner, R., & Reamer, J. (1920). A mental and educational survey of schools for the deaf. *American Annals of the Deaf, 65*, 451-472.
- Prifitera, A., & Dersh, J. (1993). Base rates of WISC-III diagnostic subtest patterns among normal, learning-disabled, and ADHD samples. *Journal of Psychoeducational Assessment, WISC-III Monograph Series*, 43-55.
- Prifitera, A., Weiss, L. G., Saklofske, D. H., & Rolfhus, E. (2005). The WISC-IV in the clinical assessment context. In A. Prifitera, D.H. Saklofske, & L.G. Weiss (Eds.), *WISC-IV clinical use and interpretation* (pp. 3-32). Burlington, MA: Elsevier Academic Press.
- Quigley, S., & Kretschmer, R. E. (1982). *The educational of deaf children*. Baltimore, MD: University Park Press.

- Quigley, S., & Paul, P. (1991). Cognition and Language. In S. Gregory & G. M. Hartley (Eds.), *Constructing deafness* (pp. 55-60). London, UK: Printers Publishers Limited.
- Raju, N.S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*, 517-529.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552-566.
- Reschly, D. J., & Grimes, J. P. (2002). Best practices in intellectual assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV: Vol. 2.* (pp. 1337-1350). Maryland: National Association of School Psychologists.
- Rettenback, R., Dillar, G., & Sireteneau, R. (1999). Do deaf people see better? Texture segmentation and visual search compensate in adult but not in juvenile subjects. *Journal of Cognitive Neuroscience, 11*, 560-583.
- Richardson, J. T.E., McLeod-Gallinger, J., McKee, B.G., & Long, G. L. (2000). Approaches to studying in deaf and hearing students in higher education. *Journal of Deaf Studies and Deaf Education, 5*, 156-173.
- Roid, G.H. (2003). *Stanford Binet Intelligence Scales, Fifth Edition, examiner's manual.* Itasca, IL : Riverside Publishing.
- Rosenstein, J. (1961). Perception, cognition and language in children. *Exceptional Children, 27*, 176-284.

- Saklofske, D. H., Rolfhus, E., Prifitera, A., Zhu, J., & Weiss, L. G. (2005). Clinical Interpretation of the WISC-IV FSIQ and GIA. In A. Prifitera, D.H. Saklofske, & L.G. Weiss (Eds.), *WISC-IV clinical use and interpretation* (pp. 33-65). Burlington, MA: Elsevier Academic Press.
- Sattler, J. M., & Hardy-Braz, S. T. (2002). Hearing impairments. In J. M. Sattler (Ed.), *Assessment of children: Behavioral and clinical applications* (4th ed., pp. 377-389). La Mesa, CA: Jerome M. Sattler Publisher, Inc.
- Schaubroeck, J., & Green, S. G. (1989). Confirmatory factor analytic procedures for assessing change during organizational entry. *Journal of Applied Psychology, 74*, 892-900.
- Schmitt, N. (1982). The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research, 17*, 343-358.
- Schwean, V.L., & Saklofske, D.H. (2005). Assessment of Attention Deficit Hyperactivity Disorder with the WISC-IV. In A. Prifitera, D.H. Saklofske, L.G. Weiss (Eds.), *WISC-IV clinical use and interpretation* (pp. 235-280). Burlington, MA: Elsevier Academic Press.
- Schwean, V.L., Saklofske, D.H., Yackulic, R.A., & Quinn, D. (1993). WISC-III performance of ADHD children. *Journal of Psychoeducational Assessment, WISC-III Monograph Series, 56-70*.
- Slate, J. R., & Fawcett, J. (1995). Validity of the WISC-III for deaf and hard of hearing persons. *American Annals of the Deaf, 140*, 250-254.

- Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spearman, C., (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.
- Spencer, J. (1885). *The principles of psychology*. New York: S. Appelton. (Original work published 1855).
- Steenkamp, J.E.M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.
- Stern, W. (1914). *The psychological methods of testing intelligence* (Educational Psychology Monographs, No. 13; G. M. Whipple, Trans.). Baltimore: Warwick & York. (Original work published 1912).
- Sternberg, R. J., & Detterman, D. K. (1986). *What is intelligence?: Contemporary viewpoints on its nature and definition*. Norwood, NJ: Ablex.
- Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly*, 47, 1-41.
- Sullivan, P. M., & Montoya, L. A. (1997). Factor analysis of the WISC-III with deaf and hard-of-hearing children. *Psychological Assessment*, 9, 317-312.
- Sullivan, P. M., & Schulte, L. E. (1992). Factor analysis of the WISC-R with deaf and hard-of-hearing children. *Psychological Assessment*, 4, 537-540.
- Swanson, C.H. (2002). WISC-III processing speed and freedom from distractibility factors in learning disabled and ADHD populations. *Dissertation Abstracts International: Section A: Humanities and Social Sciences*, 63 (6-A), 2137.

- Tanaka, J. S. (1993). Multifaceted conceptions of fit in structure equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of the Cattell-Horn-Carroll theory and cross-age invariance of the Woodcock-Johnson tests of cognitive abilities III. *School Psychology Quarterly, 19*(1), 72-87.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.
- Thurstone, L. L. (1938). The perceptual factor. *Psychometrika, 3*, 1-17.
- Tye-Murray, N. (2004). *Foundations of aural rehabilitation: Children, adults, and their family members* (2nd ed.). Clifton Park, NY: Delmar Learning.
- Vandenberg, R. J., & Self, R. M. (1993). Assessing newcomers' changing commitment to the organization during the first 6 months of work. *Journal of Applied Psychology, 78*, 557-568.
- Vernon, M. (1967). Relationship of language to the thinking process. *Archives of General Psychiatry, 16*, 325-333.
- Wasserman, J. D., & Tulskey, D. S. (2005). A history of intelligence assessment. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 3-22). New York: Guilford Press.
- Wechsler, D. (1939). *Wechsler-Bellevue Intelligence Scale*. New York: Psychological Corporation.

- Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York: The Psychological Corporation.
- Wechsler, D. (1955). *Wechsler Adult Intelligence Scale*. New York: Psychological Corporation.
- Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence*. New York: Psychological Corporation.
- Wechsler, D. (1974a). *Selected papers of David Wechsler*. New York: Academic Press.
- Wechsler, D. (1974b). *Wechsler Intelligence Scale for Children—Revised*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corporation.
- Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence—Revised*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1997). *WAIS-III WMS-III technical manual*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence—Third Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2003). *WISC-IV technical and interpretive manual*. San Antonio, TX: The Psychological Corporation.

- Weiss, L. G., Prifitera, A., & Saklofske, D. H. (2005). Interpreting the WISC-IV index scores. In A. Prifitera, D.H. Saklofske, & L.G. Weiss (Eds.), *WISC-IV clinical use and interpretation* (pp. 71- 89). Burlington, MA: Elsevier Academic Press.
- Wilson, M., Bettger, J. G., Niculae, I., & Klima, E. S. (1997). Modality of language shapes working memory: Evidence from digit span and spatial span in ASL signers. *Journal of Deaf Studies and Deaf Education*, 2, 150-160.
- Wilson, M., & Emmorey, K. (1997a). A visuo-spatial “phonological loop” in working memory: Evidence from American Sign Language. *Memory & Cognition*, 25, 313-320.
- Wilson, M., & Emmorey, K. (1997b). Working memory for sign language: A window into the architecture of the working memory system. *Journal of Deaf Studies and Deaf Education*, 2, 121-130.
- Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment*, 8, 231-258.
- Zhu, J., & Weiss, L. (2005). The Wechsler scales. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 297-324). New York: Guilford Press.

Footnotes

¹ The American Psychological Association recommends using person first language (e.g., children who are deaf or hard-of-hearing). However, the term “deaf and hard-of-hearing children/adults/individuals” or “D/HOH children/adults/individuals” is used in this paper in order to be consistent with the way in which deaf people in North American define themselves. This usage is also consistent with that used in the current literature.

² In American Sign Language (ASL), the sign for the number nine and the letter F are the same. In conversational speech, the dual meaning of the handshape does not pose a comprehension problem because the recipient is able to use context clues to decipher the intended meaning of the sign. However, the dual meaning presents a significant problem when considering the Letter-Number Sequencing subtest. Within this task, there are no context clues that allow the examinee to figure out the meaning of the sign. Within the 30 item sets of the Letter-Number Sequencing subtest, seven of them contain either the number nine or the letter F. Knowing the meaning (i.e., whether the hand signal denotes a letter or a number) is a necessary precursor to providing a correct response on this subtest. The only way a psychologist can avoid this problem is to change the test items to eliminate the use of this synonymous sign; however, this would be a clear violation of test standardization. For this reason and the reasons mentioned above, many psychologists do not give the Letter-Number Sequencing subtest to D/HOH children (N. Kordus, personal communication, May 21, 2011).

Appendixes

Appendix A

Description of the WISC-IV Indexes

The VCI is comprised of subtests intended to measure comprehension, reasoning and conceptualization (Wechsler, 2003). The VCI composite was modified to place a greater emphasis on reasoning and comprehension and rely less on acquired knowledge (i.e., crystallized knowledge) than its predecessor, the VIQ. However, verbal reasoning always requires some degree of acquired knowledge (e.g., articulating how two words are alike requires prior knowledge of each of the words). Compared to the WISC-III VIQ, the VCI is less confounded with other cognitive functions (e.g., working memory) and measures a narrower domain of cognitive functioning. As a result, the VCI is thought to be a more pure measure of verbal reasoning (Zhu & Weiss, 2005). All the subtests of the VCI (i.e., Similarities, Vocabulary, Comprehension, Information and Word Reasoning) load onto the CHC broad ability G_c (Alfonso, Flanagan & Radwan, 2005).

The WISC-IV PRI intends to invoke more nonverbal reasoning and less visual spatial skills than the WISC-III PIQ (Weiss, Prifitera & Saklofske, 2005) and is less confounded with processing speed (Zhu & Weiss, 2005). Three new PRI subtests (i.e., Matrix Reasoning, Picture Concepts and Word Reasoning) replaced three old WISC-III performance subtests (i.e., Mazes, Picture Arrangement and Object Assembly). The three new PRI subtests are thought to provide a more accurate measure of fluid reasoning and better assess the ability to reason with novel information (Prifitera, Weiss, Saklofske & Rolfhus, 2005). Two of the PRI subtests (i.e., Matrix Reasoning and Picture Concepts) load onto the CHC broad ability G_f , while the other two subtests (i.e., Block Design and Picture Completion) load onto the CHC broad ability G_v (Alfonso, Flanagan & Radwan, 2005). Despite these radical changes, the WISC-IV developers recommend that the VCI and PRI be substituted for the VIQ and PIQ, respectively, in clinical interpretations and evaluations (Prifitera, Weiss, Saklofske & Rolfhus, 2005).

The PSI is comprised of subtests intended to measure the speed of graphomotor and mental processing (Wechsler, 2003). The authors of the WISC-IV claim that the PSI subtests directly measure speed and accuracy, as well as the ability to scan and track simple visual

information. Performance on the PSI subtests reveals the rapidity with which an examinee can process simple information without making errors (Weiss, Prifitera & Saklofske, 2005; Zhu & Weiss, 2005). The three PSI subtests (i.e., Symbol Search, Coding and Cancellation) load onto the CHC broad ability Gs (Alfonso, Flanagan & Radwan, 2005).

Appendix B

Terms Unique to Deafness

Degree of Hearing Loss

A hearing loss is defined as the level of intensity (measured in decibels [dB]) needed for a person to perceive sound at a specific frequency (measured in Hertz [Hz] or cycles per second) (Sattler & Hardy-Braz, 2002). When measuring hearing loss, the most important frequencies to assess are those between 500 and 2,000 Hz, the frequency range of most speech sounds. Therefore, when diagnosing hearing loss, the major concern becomes what intensity is needed to hear frequencies in the speech range (Tye-Murray, 2004). There are four common classifications used to describe the degree of hearing loss: (a) mild (21-40 dB), (b) moderate (41-60 dB), (c) severe (61-90 dB), and (d) profound (> 91 dB) (Northern & Downs, 2002). Another way to describe the severity of a hearing loss is Pure Tone Average (PTA), the average intensity ratings (measured in dB) needed for an individual to hear frequencies of 500 Hz, 1000 Hz and 2,000 Hz (Tye-Murray, 2004). When describing hearing loss, it is also important to explain whether the impairment is unilateral (i.e., present in one ear) or bilateral (i.e., present in both ears) (Northern & Downs, 2002).

Type of Hearing Impairment

There are three main types of hearing impairments: (a) conductive impairments, (b) sensorineural impairments and (c) mixed impairments. Conductive impairments are those in which the problem transmitting sound lies in the middle or outer ear (e.g., blockage of the auditory canal, fluid in the ear, ruptured eardrum, and calcification of the bones in the middle ear). Sensorineural impairments are the result of damage to the cochlea (i.e., the organ that converts sound waves to neural impulses) or auditory nerve (i.e., the nerve connecting the cochlea to the brain). Mixed impairments are the result of both conductive and sensorineural impairments (Northern & Downs, 2002).

Age of Onset

Although doctors typically diagnose hearing loss onset as either congenital (i.e., a hearing loss present at birth) or adventitious (i.e., a hearing loss acquired after birth), most psychologists make the distinction between prelingual onset (i.e., a hearing loss present

before the acquisition of oral language) and postlingual onset (i.e., a hearing loss that occurs after the acquisition of oral language) (Braden, 2000; Tye-Murray, 2004). However, not all psychologists agree on the age at which oral language skills are “acquired.” Some psychologists argue that children as young as two years have acquired oral language because most children this age can utter one-word sentences. Other psychologists, however, say children do not acquire oral language skills until the age of five, at which point most children understand and use basic grammar, syntax and coherent conversational skills (Braden, 2000).

Deaf vs. Hard-of-Hearing

The labels of deaf and hard-of-hearing can vary depending on the context (e.g., medical vs. educational). Medically, the distinction between deaf and hard-of-hearing is based on the severity of the hearing loss. Typically, hard-of-hearing is defined as a hearing loss between 20 and 70 dB (in the mild to moderate classification range) whereas a diagnosis of deafness requires a hearing loss greater than 70 dB (in the severe to profound classification range) (Tye-Murray, 2004).

However, medical diagnoses can differ from functional diagnoses. Whereas medical diagnoses revolve around numbers (e.g., PTA) and specific anatomical abnormalities, functional diagnoses of deaf or hard-of-hearing are established on the ability to acquire and use oral language. In a functional sense, the distinction between deaf and hard-of-hearing is the ability to comprehend and produce oral speech (with or without amplification) (J.P. Braden, personal communication, September 11, 2006). Hard-of-hearing individuals can generally respond to speech and other auditory stimuli, whereas deaf individuals usually cannot understand speech or other sounds (National Dissemination Center for Children with Disabilities, 2004). There is an imperfect relationship between medical and functional definitions of hearing loss. For example, an individual with a hearing loss of 80 dB is medically diagnosed as deaf; however, with the use of hearing aids the individual is able to primarily communicate through speech, functionally falling within the hard-of-hearing classification.

Educational labels of hearing loss often differ from both medical and functional definitions and can vary across states. The definitions for deafness and hearing impairment proposed by the Individuals with Disabilities Education Improvement Act of 2004 (IDEIA-

04) do not use quantifiable classification criteria (e.g., a specific hearing loss in dB). A hearing impairment, as defined by IDEIA is “an impairment in hearing, whether permanent or fluctuating, that adversely affects a child’s educational performance, but that is not included under the definition of deafness...” (Council for Exceptional Children, 2005). According to IDEIA, deafness is a “hearing impairment so severe that the child is impaired in processing linguistic information through hearing, with or without amplification, that adversely affects a child’s educational performance” (Council for Exceptional Children, 2005). States can use the federal definitions or can create their own definitions to determine eligibility for special education as long as their standards meet the minimum requirements established in the federal definitions. For this reason, educational terminology and special education eligibility requirements vary across states. For example, a student with a hearing loss of 40 dB might be eligible to receive special education in one state, but not in a neighboring state. In some states, a student can receive special education for a unilateral hearing loss and in other states the same student would be denied similar services (Bienenstock & Vernon, 1994).

The differences in the definitions of deaf and hard-of-hearing from context to context and state to state make conducting research with this special population more challenging. For this reason, it is imperative that researchers collect demographic data regarding the characteristics of the hearing loss and report these findings to the consumers.

Prevalence Rates

According to the data collected by the U.S. Department of Education, Office of Special Education Programs, in the fall of 2006 over 80,000 children between the ages of 3 and 21 received services under the category of “hearing impairments” which includes deafness. In other words, approximately 1.2% of the children receiving special education services in 2006 had a primary diagnosis of deaf or hard-of-hearing (Individuals with Disabilities Education Act Data, 2006). However, this count is likely an underestimate of the total number of children with hearing loss due to the fact that some deaf and hard-of-hearing children may qualify for special education under another category and others may be served exclusively in regular education classrooms (National Center on Birth Defects and Developmental Disabilities, 2004).

Appendix C

Newsletter Posting to Listservs

The WISC-IV with Deaf and Hard-of-Hearing Students

Jeffery P. Braden, Ph.D.

Hailey Krouse, M.S.

NC State University

The Wechsler Intelligence Scale for Children--Fourth Edition (WISC-IV) is a newly updated version of the most popular intelligence test used with deaf and hard-of-hearing (DHOH) students in the US. Unfortunately, although the technical data presented with the WISC-IV describes the instrument's characteristics with many special populations, no data are provided for DHOH students.

We invite psychologists and other professionals, who are members of national interest groups focusing on DHOH students, to provide me with anonymous archival data describing the performance of DHOH students on the WISC-IV. We will send participating professionals an Excel spreadsheet already formatted to receive data. Professionals will insert data, but NOT identifying information, into the spreadsheet and return it to me via email for analysis. Although the scope of such a study is typically beyond any one professional or entity, a collaborative effort could produce a large and varied sample of DHOH students. Note that, because the study will collect anonymous, archival data, it will NOT be necessary to obtain permission from individual students or their parents/guardians for participation.

This information will be useful to professionals (e.g., psychologists, speech therapists) who use the WISC-IV with DHOH populations, and will indirectly benefit DHOH students by providing these professionals with a better understanding of the strengths and limitations of the WISC-IV when used with DHOH populations. Please contact me (hekrouse@ncsu.edu) to participate in this project.

Thanks!

Hailey Krouse, M.S.
School Psychology Doctoral Student
North Carolina State University

Appendix D

Posting for Websites

We are seeking data on the clinical use of the WISC-IV with deaf and hard-of-hearing (DHOH) populations. We are posting this announcement on this Listserv because it is likely to reach qualified professionals who use the WISC-IV with DHOH populations.

The Wechsler Intelligence Scale for Children--Fourth Edition (WISC-IV) is a newly updated version of the most popular intelligence test used with DHOH students in the US. Although the technical data presented with the WISC-IV describes the instrument's characteristics with many special populations, no data are provided for DHOH students.

We invite psychologists and other professionals, who are members of national interest groups focusing on DHOH students, to provide us with anonymous archival data describing the performance of DHOH students on the WISC-IV. Although the scope of such a study is typically beyond any one professional or entity, a collaborative effort could produce a large and varied sample of DHOH students.

This information will be useful to professionals (e.g., psychologists, speech therapists) who use the WISC-IV with DHOH populations, and will indirectly benefit DHOH students by providing these professionals with a better understanding of the strengths and limitations of the WISC-IV when used with DHOH populations.

If you are interested in participating or would like to learn more about this study please view the attached consent form. If you would like to contribute to this study, please print a copy of the consent form, fill it out, and send it to the address designated at the bottom of the form. Once we have received a copy of your consent form, we will email you an Excel Spreadsheet in which we will ask you to fill in subject information and email it back to us. The North Carolina State University Institutional Review Board (IRB) has approved this study. Thank you so much for your time and contribution, this study would not be possible without your help!

Thanks!

Jeff Braden, Ph.D.
Dept. of Psychology
Box 7650
North Carolina State University
Raleigh, NC 27695-7650
Email: jeff_braden@ncsu.edu

Hailey Krouse, M.S.
School Psychology Doctoral Student
North Carolina State University
Email: hekrouse@ncsu.edu

Appendix E

Consent Form

The WISC-IV with Deaf and Hard-of-Hearing Students

Why do this study?

The WISC-IV was released without any data describing the performance of DHOH children on the scale. There are two reasons why this is a problem: (a) previous versions of the Wechsler Scales are the most popular method for assessing the intelligence of DHOH clients, and (b) the WISC-IV (and in particular, the language-reduced scale) is changed substantially from previous editions. This means that the characteristics of the new version of the Wechsler are unknown, and changes from previous editions are sufficiently substantial to draw caution to assumptions that the new edition will have the same (excellent) characteristics as its predecessors when used with DHOH students.

What do you want from me?

I would like interested professionals who already have WISC-IV data on DHOH clients to provide item, subtest, and scale data to me. By combining data across many professionals, we can overcome the problems associated with the low incidence of deafness and obtain a sufficiently large sample to conduct meaningful psychometric analyses of the WISC-IV (e.g., reliability, validity, utility). I estimate that you would need about 10 minutes per client to enter all the test and demographic data into an Excel spreadsheet. I do NOT want identifying information for clients; all data will be sent to me in an encrypted file.

What will you do with the data?

I will aggregate the data across all contributing members of the study group. Once the data are aggregated, I will conduct measurement invariance analyses. I will ask The Psychological Corporation to supply me with a matched group of normal-hearing participants to compare and contrast the test's characteristics with DHOH vs. normal-hearing children. I plan to present and publish the results of this study at relevant professional meetings, journals, etc.

Why should I participate?

There are two reasons why you should contribute data to this study: (a) professional ethics, and (b) professional recognition. With respect to ethics (e.g., AERA, APA, NCME, 1999 Standards for educational and psychological testing 3rd ed.), test users must use data from relevant populations to inform test use. When those data are lacking, the user (and the publisher) have an obligation to collect data to understand the psychometric characteristics of the test when used with special populations. Therefore, ethics suggests that we share an obligation to get data on the WISC-IV with DHOH populations, because we are using the

WISC-IV with those populations. Second, with respect to recognition, I will publish all studies under corporate authorship (Professionals Serving Deaf and Hard-of-Hearing Clients). I will list the names, titles, and institutional affiliations of all who contribute to the study in an appendix to all publications, presentations, or other dissemination of the data. If you do not want to be recognized, you may contribute data and I will list you under anonymous contributors (who will be counted, not named). There will be no other reward for participation.

What if I don't want to participate?

That's fine—don't do anything. There are no negative consequences for declining participation, and you can withdraw at any time without penalty. An electronic copy of the results of the study will be available free of charge to anybody who asks, whether or not they decide to participate. I will post the copy to the listserv host site when the analyses are complete.

OK, I'm interested. What do I do next?

Complete this letter of consent and send it to me (see below) via standard mail. If you have any questions or concerns that arise in connection with your participation in this study, you may contact Hailey Krouse at hekrouse@ncsu.edu. This study has been reviewed and approved by the NCSU Institutional Review Board (IRB). If you have questions or concerns about your rights or the rights of the DHOH students, please contact the NCSU IRB administrator, Ms. Debra Paxton, at (919) 515-4514 or debra_paxton@ncsu.edu.

I have read the above and have been given the opportunity to ask questions. I agree to participate in this research with the understanding that I may withdraw without penalty at any time. I have kept a copy of this letter for my records.

Date

Signature

Password (Please type or print. Minimum 6 characters with at least 1 digit and 1 capital letter.)

Please type or print the e-mail address to which we should send your Excel files above.

Please send a copy of your signed/dated letter with your password to:

Hailey Krouse, M.S.

Dept. of Psychology

Box 7650

North Carolina State University

Raleigh, NC 27695-7650

Appendix F

Restriction of Range Formula

$$r_{\text{corrected}} = r_{xy} (S_x/s_x) / \sqrt{(1 - r_{xy}^2 + r_{xy}^2 (S_x/s_x)^2)}$$

Where:

r_{xy} = uncorrected correlation coefficient

S_x = population standard deviation

s_x = sample standard deviation