



2007-12-07

Existing Practice and Proposed Changes in Cognitive Assessment of Utah Students Identified as Deaf and Hard Hearing

Leah Voorhies

Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Counseling Psychology Commons](#), and the [Special Education and Teaching Commons](#)

BYU ScholarsArchive Citation

Voorhies, Leah, "Existing Practice and Proposed Changes in Cognitive Assessment of Utah Students Identified as Deaf and Hard Hearing" (2007). *All Theses and Dissertations*. 1281.

<https://scholarsarchive.byu.edu/etd/1281>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

EXISTING PRACTICE AND PROPOSED CHANGES IN
COGNITIVE ASSESSMENT OF UTAH STUDENTS IDENTIFIED AS DEAF
AND HARD OF HEARING

by

Leah Voorhies

A dissertation submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Counseling Psychology and Special Education

Brigham Young University

December 2007

Copyright © 2007 Leah Voorhies

All Rights Reserved

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a dissertation submitted by

Leah Voorhies

This dissertation has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date

Melissa Allen Heath, Chair

Date

Lane Fischer

Date

Ellie L. Young

Date

Richard Sudweeks

Date

Tina T. Dyches

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the dissertation of Leah Voorhies in its final form and have found that (1) its format, citations and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Melissa Allen Heath
Chair, Graduate Committee

Accepted for the Department

Aaron P. Jackson
Graduate Program Coordinator

Accepted for the College

Barbara Culatta
Associate Dean, McKay School of Education

ABSTRACT

EXISTING PRACTICE AND PROPOSED CHANGES IN COGNITIVE ASSESSMENT OF UTAH STUDENTS IDENTIFIED AS DEAF AND HARD OF HEARING

Leah Voorhies

Department of Counseling Psychology and Special Education

Doctor of Philosophy

This study presented the past, current, and proposed practice of intelligence testing with a unique population, students identified as deaf and hard of hearing (D/HH). As a basis for describing the cognitive ability of Utah's D/HH students and to improve practice guidelines, 61 D/HH students served by Utah Schools for the Deaf and the Blind (USDB) were administered the Universal Nonverbal Intelligence Test (UNIT) standard battery and the Perceptual Reasoning Index (PRI) subtests from the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV). Based on these data, composite score distributions were described and compared with national standardization samples.

Participants' WISC-IV PRI scores are summarized with the following descriptive statistics: $M = 88.95$, 11.05 points below the standardization sample's mean; $SD = 14.55$; skew = $-.74$; and $SE = .31$. Comparing the USDB D/HH sample's WISC-IV PRI scores with the WISC-IV standardization sample's distribution of scores, the participants' scores

were significantly lower (two-tailed p -value of $<.0001$). Participants' UNIT Standard Battery Composite scores are summarized with the following descriptive statistics: $M = 90.74$, 9.26 points less than the standardization sample's mean; $SD = 13.97$; skew = $-.55$; and $SE = .31$. Comparing this sample's UNIT composite scores with the standardization sample, the participants' scores were significantly lower (two-tailed p -value of $<.0001$). Additionally, a Pearson correlation compared each participant's scores on the WISC-IV PRI with the corresponding score on the UNIT Standard Battery Composite, yielding a correlation coefficient of $.75$ with a two-tailed p -value $<.0001$.

Recommendations for future guidelines regarding cognitive assessment of Utah's D/HH students are presented. In particular, this research supported administering the UNIT rather than the WISC-IV. Though no assessment is *language free*, the UNIT's administration uses simple gestures for directions, rather than spoken language. Additionally, D/HH students were included in the standardization sample. Furthermore, administering one assessment, rather than several, consumes less time for the examiner and the student, saving money and decreasing student time away from classroom instruction.

ACKNOWLEDGMENTS

When I was nine years old my maternal grandfather told me that if I became a lawyer he would disown me. My grandfather was a psychologist. The statement baffled me as a child, but haunted me as a young adult. And, by the time I was looking at graduate programs I'm sure he'd forgotten he'd said it. But I hadn't forgotten and I did actually sort of want to go to law school. But as obeying my grandfather was always in the back of my mind, the law was out. Instead, I applied to psychology programs, thinking that couldn't possibly disappoint him. My grandfather passed away two years ago, but not before he knew and felt immense pride in that I had followed in his career path. And, I'm so glad I did. I love psychology. I love learning about and thinking about human behavior and motivation. I am grateful to my grandfather for pointing me all those many years ago in the exactly right direction.

I learned all the really good psychology stuff I know from the faculty in the Counseling Psychology and Special Education department and the Counseling Center at BYU. I am grateful to them for teaching and supporting me. I know that I love psychology in part, because of the examples of those who presented it to me. They are inspired, dedicated and caring teachers and therapists who serve their students, me, and the university with integrity. They represent all that is good in the field. I'm especially grateful to Dr. Melissa A. Heath, because she contributed immensely to this paper, and kept right by me throughout the process.

I owe a great professional debt to my mentors and friends at the Utah Schools for the Deaf and the Blind. They are lovely every one, life-long educators, administrators and

school psychologists alike. Especially Tami Gear, who sat next to me for hours on end, entering data and editing text.

And, then there's my sweet family. My extended family who adored me and just loved to be "analyzed" and "therapized." My brothers who often gave me grief about being single and also about choosing psychology and education over something that made real money. My sister who was a genuine delight and the only person who thought all my jokes were funny and danced with me in movie theatres. My father who encouraged/reminded me often to finish this paper and who is my most steadfast and vocal fan. And, my sweet, English-loving/teaching mother, she's simply an angel. She reminded time and again of another thing my grandfather told me, "Don't let school interfere with your education." I took that statement too to heart, and I have loved the schooling I've received, loved the education I've received. I'm one lucky, happy, grateful psychologist.

TABLE OF CONTENTS

ABSTRACT	v
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS	ix
LIST OF TABLES	xii
LIST OF FIGURES	xiii
Introduction.....	1
Assessing D/HH with Traditional Tests of Cognitive Ability.....	2
Description of D/HH Population.....	3
Causes of Hearing Loss.....	4
Demographics of D/HH Students.....	4
Standardized Assessment of Intellectual Ability.....	5
Wechsler Intelligence Scale for Children-Fourth Edition	6
Accommodations for Administering the WISC-IV to D/HH.....	7
Description of Nonverbal Intelligence Tests.....	9
USDB’s Standard Procedure for Cognitive Assessment.....	10
Statement of Purpose.....	13
Literature Review.....	15
Cognitive Assessment of D/HH	15
Overview of Assessment Research	18
Wechsler Performance Scales.....	20
Non-Verbal Tests.....	21
Key Research Findings Regarding IQ Assessment of D/HH	21

Validity of IQ Testing with D/HH	22
Factor Analysis of IQ Assessment Scores of D/HH	29
Item Analysis of D/HH Students' Responses on the WISC and UNIT	31
Subcategories of D/HH: Implications for Interpreting IQ Scores	34
Form of Communication Utilized by Student	35
Type of Educational Program	36
Conclusions and Future Directions.....	36
Research Questions.....	37
Methods.....	39
Selection of Participants.....	39
Sample Demographics.....	40
Setting.....	40
Instruments.....	40
UNIT.....	40
UNIT Standardization Sample.....	42
UNIT Standard Battery.....	42
Reliability of UNIT.....	43
Validity of UNIT.....	44
WISC-IV.....	46
WISC-IV Standardization Sample.....	46
Reliability of WISC-IV.....	47
Validity of WISC-IV.....	48
Procedures.....	50

Data Analysis.....	51
Results.....	53
Research Question One.....	53
Research Question Two.....	54
Research Question Three.....	55
Research Question Four.....	56
Research Question Five.....	56
Discussion.....	58
Limitations and Future Research.....	62
Conclusions and Recommendations.....	63
References.....	65
APPENDIX A.....	72
APPENDIX B.....	77

LIST OF TABLES

1. D/HH Students' Number of WISC-III Subtest Items Exhibiting Differential Item Functioning	34
2. Participant Demographics.....	41
3. UNIT Reliability Coefficients	44
4. UNIT External Validity Coefficients.....	45
5. WISC-IV Indices' Reliability Coefficients.....	48
6. WISC-IV Indices' Internal Consistency Coefficients	49
7. Descriptive Statistics of Composite Scores for the WISC-IV PRI and UNIT Standard Battery	53

LIST OF FIGURES

1. WISC-IV PRI Score Distribution.....	54
2. UNIT Standard Battery Composite Score Distribution.....	55
3. Scatter Plot of UNIT and WISC-IV PRI Scores.....	57

Introduction

In the United States 10,000,000 individuals are identified as hard of hearing and 1,000,000 are identified as functionally deaf (Mitchell, in press). About 4% of the functionally deaf individuals are under 18 years of age (Mitchell). In regard to accommodating for educational needs, a major debate focuses on the validity and utility of traditional intelligence tests to measure cognitive abilities of D/HH students. In 2005, Braden concisely summarized the reason for this debate: “Intellectual assessment of clients who are hard-of-hearing or who are deaf is difficult yet essential” (p. 351).

The difficulty centers on the fact that hearing loss impairs *all* facets of a child’s development, including the child’s ability to access and develop language, thereby hindering the child’s ability to develop emotional foundations and critical relationships with parents and caretakers. Poorly developed and even delayed language development then hinders social relationships with peers. In addition to impaired or delayed social development, as the child enters school, educational deficits and difficulty learning in conventional classrooms becomes apparent, necessitating educational accommodations (Braden, 2005).

Although controversy surrounds the intellectual assessment of D/HH students, acquiring accurate information about the abilities of D/HH individuals is essential, “a vital component for planning educational, social, vocational, and even medical interventions” (Braden, 2005, p. 351). Thus, the major conundrum for examiners is accurately interpreting assessment data, more specifically, “differentially diagnosing intellectual deficits from experiential deficits” (Braden, p. 352).

Assessing D/HH with Traditional Tests of Cognitive Ability

Unfortunately, with D/HH students it is nearly impossible to accurately differentiate cognitive deficits using traditional measures of intelligence (Braden & Athanasiou, 2005). Traditional tests are administered verbally, an inherent disadvantage to the D/HH. Because hearing loss typically results in delayed language development, D/HH individuals have a difficult time understanding overall directions and information related to specific test items. They may also struggle to respond in an age-appropriate manner. Also, during test administration, the instructions relayed to the student do not easily translate into American Sign Language. Therefore, students who communicate with sign language cannot adequately access the information they need in order to respond appropriately (Braden, 2005).

Another reason for the debate about using traditional intelligence tests with the D/HH is that such tests are not specifically standardized on a population of D/HH students, nor include D/HH students in their standardization sample. And, as stated earlier, the test's standardized administration must be modified for D/HH students, especially for those communicating with sign language. In particular, the instructions must include information the student understands, clearly delineating what is required and how to communicate a response. These issues compromise the validity of the students' scores (Maller, 2003). However, because scores associated with cognitive abilities are necessary, these standardized tests continue to be used, regardless of the questionable validity of the obtained scores.

Unfortunately, because deafness is a low incidence disability (U.S. Department of Education, 2002), it is not cost-effective for researchers to create or publishers to invest in a test standardized solely on the D/HH population. At present, no such intelligence test

exists. Maller emphasized, “. . . countless deaf examinees regularly are evaluated with tests that lack necessary psychometric evidence” (2005, p. 1095).

Hence, regardless of the compromised validity of scores and the questionable reliance on traditional intelligence tests, these measures have been and continue to be administered to the D/HH. Two of the most common intelligence tests are the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV; Wechsler, 2003) and the Stanford-Binet – Fifth Edition (SB-V; Roid, 2003). Although these assessment instruments have subtests that appear to be nonverbal, as they do not require the child to respond verbally, the directions are often complicated and must be administered using language, whether verbal or signed. Additionally, directions and responses require D/HH students to negotiate an internal translation, similar to translating from one language to another. This requires extra time and poses the uncertainty of whether the student truly understood what was expected. Bottom line, the WISC-IV and SB-V subtests cannot be considered as *non-verbal*, but are more accurately considered as *language reduced* (McCallum, Bracken, & Wasserman, 2001).

Description of D/HH Population

There are generally four different terms used to describe individuals with a hearing loss. Those who use American Sign Language (ASL) to communicate and share a common culture, no matter the degree of their hearing loss, choose to be identified as *Deaf*, with the capitalized “D” representing their involvement in the Deaf community. Those who have a severe to profound hearing loss, meaning they cannot hear speech without amplification, but who use spoken English to communicate, are considered *deaf*, with a lowercase “d.” Individuals who have a mild to moderate hearing loss and who use spoken English to communicate are referred to as *hard of hearing*. The fourth term,

hearing impaired, is no longer a politically correct term to refer to individuals with a hearing loss. However, based on the 2004 Individuals with Disabilities Education Act, the educational classification for students with a hearing loss, who are not deaf, is still hearing impaired. For the purpose of this study, all children with an educational classification of *deaf* or *hearing impaired* will be identified as D/HH, currently the most politically correct and universally used abbreviation describing the entire population of individuals with hearing loss (Klein & Parker, 2002).

Causes of Hearing Loss

The National Institute on Deafness and Other Communication Disorders (<http://www.nidcd.nih.gov>) reports that in 2004 approximately 1,055,000 children under the age of 18 in America were identified with hearing loss. There are many different causes of hearing loss, including in utero rubella, perinatal cytomegalovirus, meningitis, Usher's and CHARGE syndromes, and side effects of large doses of erythromycin to treat infections. Although hearing loss can be an inherited condition, less than 10% of D/HH children are born to D/HH parents (Brown, 1986; Gallaudet Research Institute, 2003). Furthermore, hearing loss is also related to and associated with other disabilities. In fact, approximately 40% of D/HH individuals have concomitant disorders, including learning disabilities, mental retardation, CHARGE syndrome, and psychiatric disorders, etc. (Karchmer & Mitchell, 2003; Knoors & Vervloed, 2003).

Demographics of D/HH Students

Much of the demographic information describing the academic features of D/HH students in American comes from an annual survey by Gallaudet University. (Note: USDB participates in this survey each year, so the students whose scores are included the present study are represented in the Gallaudet survey.) The most recent widely published

Annual Survey described the 2000-2001 school year in which approximately 21,500 or 60% of the D/HH students, ages 6-21, in the United States were included (Gallaudet Research Institute, 2003). Two-thirds of those students used personal hearing aids and about five percent had cochlear implants. Approximately 50% of the students relied on communication methods other than speech (Gallaudet Research Institute).

Most of those 21,500 students who use some form of signed language instead of speech as their primary means of communication consider themselves culturally Deaf, and participate at some level with the Deaf community. As stated above, less than 10% of D/HH children are born to hearing parents (Brown, 1986; Gallaudet Research Institute, 2003). This means that only ten percent of the students in America who use sign language to communicate are born to parents who also use sign language to communicate. In fact, 72% of their family members do not sign regularly, and few sign fluently (Gallaudet Research Institute). Ultimately, this means that for 70% or more of students who use sign language to communicate, possibly the *only* interaction they have with other individuals who use sign language is at school, making educational programming of utmost importance to these students.

Standardized Assessment of Intellectual Ability

Modern day intelligence testing in the United States dates back to Alfred Binet and Theodore Simon in France (Wasserman & Tulskey, 2005). Although their initial efforts went through several revisions, in 1905 the first version contained 30 items scored with a pass/fail. Primarily, the assessment's purpose was to identify children who would not benefit from traditional school settings. Taking Binet's work, translating and modifying questions to Americanize the test, Henry Goddard then promoted intelligence testing with children in the United States. However, testing with adults, not children, was

responsible for putting intelligence testing on the fast track, ensuring a premier position of power for years to come. The American Psychological Association was heavily involved in Army testing during World War I, identifying which individuals would best perform certain duties (Wasserman & Tulsky).

Wechsler Intelligence Scale for Children-Fourth Edition

In school settings, the Wechsler intelligence tests have enjoyed a long reign of prominence since the 1960's (Wasserman & Tulsky, 2005). The current version of this series is the Wechsler Intelligence Scale for Children-Fourth Edition, WISC-IV. This test is an individually administered clinical instrument for assessing the intellectual ability of children ages 6 years through 16 years, 11 months. The WISC-IV consists of several subtests, each measuring a somewhat different aspect of intelligence. The individual's performance on these various measures is summarized in four Composite scores: (a) Verbal Comprehension (VCI), (b) Perceptual Reasoning (PRI), (c) Working Memory (WMI), and (d) Processing Speed (PSI). The Full Scale IQ (FSIQ) is one major score representing an individual's overall cognitive ability.

The VCI and the WMI are composed of subtests that require verbal responses. Overall, the majority of these subtests cannot be adequately translated into ASL, making it very complicated to administer the subtests to students using ASL as their primary means of communication. The VCI can be broken down into four standard subtests and one optional subtest: (a) Similarities is a measure of abstract reasoning ability; (b) Vocabulary is a measure of word knowledge, learning ability, fund of information, richness of ideas, memory, concept formation and language development; (c) Comprehension is a measure of the ability to express practical social knowledge and judgment; (d) Information is a measure of long-term retention of information learned at

school and at home; and (e) Word Reasoning is an optional subtest which is a measure of the ability to synthesize and analyze different types of verbal information to generate alternative concepts (Wechsler, 2003, pp. 14-15).

The WMI can be broken down into three subtests: (a) Digit Span is a measure of short-term auditory memory function; (b) Letter-Number Sequencing is a measure of short-term auditory memory and sequencing, mental manipulation, attention, visuospatial imaging and processing speed; and (c) Arithmetic is a measure of the ability to calculate and perform simple mental computation (Wechsler, 2003, pp. 14-17).

Though the subtests on the PRI and PSI have verbal instructions, the tasks do not require verbal responses. The subtest instructions can be administered using ASL but do not have a gestural administration method. This makes the tasks *language-reduced*, not nonverbal. The PRI can be broken down into four subtests: (a) Block Design is a measure of the ability to analyze and reproduce an abstract design; (b) Picture Concepts is a measure of abstract categorical reasoning ability; (c) Matrix Reasoning is a measure of visual information processing and abstract reasoning skills; and (d) Picture Completion is a measure of the ability to separate essential and nonessential detail.

The PSI can be broken down into three subtests, all of which require motor skills: (a) Coding is a measure of the ability to learn and memorize new nonverbal material while drawing it; (b) Symbol Search is a measure of visual recognition and scanning; and (c) Cancellation is a measure of processing speed, visual selective attention, vigilance and visual neglect (Wechsler, 2003, pp. 15-18).

Accommodations for Administering the WISC-IV to D/HH

The WISC-IV Administration and Scoring Manual provides several pages of guidelines for administering subtests to D/HH children. These guidelines begin by

stating, “To obtain reliable, valid, and clinically useful results, examiners must accommodate the child while minimizing modifications to standard administration procedure” (Wechsler, 2003, p. 13). To accommodate the child, the examiner must administer the test in the child’s preferred communication mode (i.e, oral, sign language and cued speech) and then administer only those subtests that the primary consultant recommended appropriate for children using each communication mode.

D/HH children who use sign language are recommended to receive the Perceptual Reasoning subtests, and then Processing Speed subtests with some caution given in interpretation. On the other hand, D/HH children who are oral can be administered the Verbal Comprehension and Perceptual Reasoning subtests, and the Processing Speed and the Working Memory subtests can be administered, using caution when interpreting scores (p. 15). The guidelines remind examiners that most examiners will resort to pointing and using gestures when administering the WISC-IV to D/HH children. However, even though this practice is permitted by the standardized instructions, Blennerhassett & Traxler (1999) reported that such practices were unclear and created confusion for many D/HH children.

Maller (2005) reviewed the WISC-IV and predicted that the WISC-IV would follow in the footsteps of previous test editions and “would continue to hold its place in history as a dominant force among individually administered intelligence tests” (p. 1095). However, she expressed concern regarding the use of the test with D/HH individuals, “Although a consultant and ‘experienced specialists’ provided judgments regarding subtests appropriateness, D/HH children were excluded from validity studies. Thus, judgments were not based on empirical research. . .” (p. 1095). Braden (2005) shared a similar concern in that the WISC-IV manual’s recommendations for subtest use with the

D/HH, “are based on logical and subjective analyses, and not on direct evidence” (p. 363).

Description of Nonverbal Intelligence Tests

An alternative to verbal measures of intellectual ability like the WISC-IV is the use of a nonverbal test. Truly nonverbal intelligence tests do not use verbal instructions, but instead utilize gestures and demonstration methods of explaining tasks. There are several nonverbal intelligence tests on the market which are designed to provide a fair assessment of intelligence for children who would be disadvantaged by traditionally language-loaded tests. Such children include the D/HH, but also children with communication disorders and those who come from non-English speaking families. These tests include: (a) the Universal Nonverbal Intelligence Test (UNIT; Bracken & MacCullum, 1998), (b) the Leiter International Performance Scale-Revised (Leiter-R; Roid & Miller, 1997), (c) the Test of Nonverbal Intelligence-Third Edition (TONI-III; Brown, Sherbenou, & Johnson, 1997), and (d) the Comprehensive Test of Nonverbal Intelligence (C-TONI; Hammill, Pearson, & Wiederholt, 1996).

The UNIT is an individually administered clinical instrument for assessing the intellectual ability of children ages 5 through 17 years old. The UNIT is a nonverbal test designed to provide a fair assessment of intelligence for children who would be disadvantaged by traditional language loaded tests. The UNIT consists of several subtests, each measuring a somewhat different aspect of intelligence. The student’s performance on these various measures is summarized in five Composite scores: (a) Memory Quotient, (b) Reasoning Quotient, (c) Symbolic Quotient, (d) Nonsymbolic Quotient, and (e) Full Scale IQ. Scores in these areas provide estimates of the individual’s intellectual abilities.

The UNIT can be broken down into six subtests. The first four subtests comprise the Standard Battery of the UNIT. Symbolic Memory is primarily a measure of short-term visual memory and complex sequential memory for meaningful material. Cube Design is primarily a measure of visual-spatial reasoning. Spatial Memory is primarily a measure of short-term visual memory for abstract material. Analogic Reasoning is primarily a measure of symbolic reasoning. The Extended Battery includes two extra subtests. Object Memory is primarily a measure of short-term recognition and recall of meaningful symbolic material. The Mazes subtest is primarily a measure of reasoning and planning behavior (Bracken & McCallum, 1998, pp. 210-219).

A description of a group comparison study with the D/HH was included in the UNIT examiner's manual. A sample of 60 female and 46 male students selected from a school for the D/HH were administered the UNIT. On the Standard Battery of the UNIT, the D/HH students received a mean Composite score difference of 6.20 points lower than a sample of hearing students matched for age, sex, race, ethnicity and parent educational level from the standardization sample (Bracken & McCallum, 1998, pp. 192-193).

USDB's Standard Procedure for Cognitive Assessment

Many schools for the D/HH across the United States, including the Utah Schools for the Deaf and the Blind (USDB), have chosen to give D/HH students multiple intelligence measures so that a range of scores or an average score can be established and used for educational programming. And, though the 2004 Reauthorization of the Individuals with Disabilities Education Act does not require that a student be administered a cognitive measure to be eligible for special education under the classification of Deaf or Hearing Impaired, the scores from cognitive measures are important to teachers and administrators for two reasons. First students who are found to

have very high (>130) and very low (<60) intelligence scores, are often placed in their home districts in either an inclusive or cluster setting based on availability and need, instead of a USDB self-contained class. And second, teachers use the scores to compare the students in their classes to one another in an effort to differentiate instruction based on ability level and similarly, to judge how hard to push them academically.

Historically, USDB used the Wechsler scales and the current version of the Test of Nonverbal Intelligence (TONI) to assess the intelligence of its students. This practice was based on the dissertation research of Cash (1994). She found that the TONI-2 scores were very similar to the scores D/HH students received when the deaf norms of the WISC-R were used. As a result of that study, USDB attempted to only give D/HH students a TONI. However, in doing so, USDB heard two significant criticisms.

First, because Utah schools have a long established history of administering the Wechsler scales, school district personnel consistently request it because they are comfortable with it, its subtests, and its interpretation. Even when school district personnel recognize that students identified as D/HH may be disadvantaged by a verbally administered intelligence test, they continue to argue that an intelligence test score is needed, and the Wechsler scales are *better than nothing*.

And, probably more importantly, Utah's state social services agencies have relied and continue to rely upon Wechsler scores to identify disabilities and subsequently the eligibility for benefits and services. Though individuals may be eligible for social security benefits based solely on their hearing loss, individuals must also have a cognitive disability to be eligible for services through the Division of Service for People with Disabilities (DSPD). These services included respite care, group housing, sheltered workshops, job coaches, and financial assistance, etc. D/HH children and/or their families

can apply for DSPD services as soon as they students are identified as having both a hearing loss and a cognitive impairment. Therefore, most parents who believe that their child qualifies for services, request that a cognitive be administered each time the student's special education eligibility must be reviewed.

When USDB personnel have attempted to present scores from an intelligence test other than a Wechsler to DSPD, the student is either turned down immediately for benefits/services, or the parents are asked to obtain additional assessment from a private provider. Additional assessment is costly and significantly delays the process of determining eligibility.

Unfortunately for D/HH individuals, neither school districts nor social service agencies appear to be concerned about the validity of WISC test scores, but rather, their concern appears to be focused on merely getting a WISC score. Neither group appears to understand that WISC test scores, that they are requesting, are based on a standardized test which did not include D/HH individuals in the standardization sample nor in the validity studies.

As such, these agencies are depending on scores with no established standard or basis of validity. Thus, the practice of administering two intelligence tests, a Wechsler scale and a nonverbal measure, has endured and continues to be the accepted practice at USDB.

Over the past three years, the USDB school psychologists have begun using the UNIT, instead of a TONI, in concert with the current version of Wechsler test, the WISC-IV. The school changed from the TONI to the UNIT because the UNIT has both a breadth of assessment (as opposed to the TONI's single factor), as well as a symbolic versus a non-symbolic scale. Further, the UNIT replaced the TONI at USDB because the

UNIT's authors considered the D/HH population when designing the test, and included D/HH children in the standardization sample. Additionally, the UNIT provides a comparison study of D/HH children to hearing children in the examiner's manual (Bracken & McCallum, 1998).

USDB's historical and current cognitive assessment practice is evidence of the debate about using traditional intelligence tests with D/HH individuals. Special education eligibility, and thus school district personnel, as well as social services eligibility in Utah require a score to represent the cognitive ability of the D/HH individual, never considering the issue Braden (2005) raised, that test scores may instead represent the D/HH individual's experiential ability. Furthermore, special educators and social servants have shown more interest in receiving a score that has historical and comfort value rather than statistical validity, actually measuring what it purports to measure. Proposing and investigating a new protocol for intellectual assessment of D/HH is in order, one that makes sense to all interested parties: special educators, social servants, and school psychologists who administer tests to D/HH individuals in USDB.

Statement of Purpose

Administering several different cognitive measures takes students out of the classroom for extended periods of instruction time, as administering the Standard Battery of the UNIT takes about 20 to 30 minutes and administering the Perceptual Reasoning Index of the WISC-IV takes about 15-25 minutes. Additionally, extra testing requires extensive professional time, already a very scarce commodity, particularly for school psychologists who have expertise in working with D/HH population. Statistically establishing that the typically administered measures are similar or that one measure

describes D/HH students better than another, would alleviate the need to administer multiple measures, saving money and saving time for both student and examiner.

The purpose of this study was threefold. First, this study proposed to describe the distribution of cognitive scores of the D/HH students currently being served by USDB. This study will provide a context for better understanding an individual D/HH student's score within the context of the USDB D/HH WISC-IV and UNIT scores, by describing statistically the cognitive scores of D/HH students at USDB. Simply stated, this study will answer the question, "what does a single USDB D/HH IQ test score mean in the context of the total sample of USDB D/HH scores on that particular test?"

Second, in conducting required intellectual assessment for students identified as D/HH (portions of the UNIT and WISC-IV as previously described), this study will determine if testing time can be reduced (by eliminating WISC-IV assessment), yet the reliability of assessment results maintained.

And finally, in regard to intellectual assessment of USDB D/HH students, dependent on research question two, the study will provide DSPD, other state agencies, and Utah school psychologists with data and rationale to either support or not support the possible use of the UNIT testing scores rather than the combination of WISC-IV *and* UNIT scores.

Literature Review

The assessment of cognitive abilities or *intelligence* has been a staple in education for decades (Fagan, 2002; Flanagan & Harrison, 2005). Intelligence tests provide a method for determining if students are gifted or cognitively disabled. Intelligence is used as a measure of students' academic potential, those having lower scores not being expected to achieve as much as those having higher scores. Historically, cognitive assessment has served as a gate keeping mechanism defining which students receive special education services, as those students who have a significant discrepancy between intelligence and achievement often benefit from and/or require the specialized instruction of special education programs/placements (Mather & Wendling, 2005).

Cognitive Assessment of D/HH

Measures of intelligence have similarly been used with students identified as D/HH (Braden & Athanasiou, 2005). Initially, researchers found a significant difference between the intelligence quotient scores of the two groups of individuals, hearing and D/HH, leading to the conclusion that the D/HH were “cognitively inferior” to their hearing peers (Pinter & Patterson, 1917).

Myklebust (1953, 1960) refuted Pinter and Patterson's conclusion by asserting that there were instead, qualitative differences in “perceptions, imagery, symbols and concepts” (p. 229) between the D/HH and hearing individuals. Similarly, Vernon (1967) examined prior studies and found that D/HH children did “remarkably well” compared to hearing peers when one considered the lack of language stimulation they experienced, and the high rates of comorbid disabilities occurring in the population (Moores, 1996). Further investigation revealed some evidence that D/HH individuals performed as well as

their hearing peers on intelligence tests with nonverbal tasks; however, the results were certainly not conclusive (Marschark et al., 2002).

Vernon's (1967) examination illustrates a long standing history of heavy reliance on verbal ability to indicate level of intelligence. Braden (1994) explained the phenomenon. "Research on intelligence with normal-hearing people uses verbal tests quite frequently because they are less expensive to purchase and administer, are often more reliable, and are better predictors of academic success than nonverbal tests" (p. 77).

However, Braden (1992) also explained that the use of verbal measures of intelligence was misleading when used with the D/HH because,

. . . the low score of a deaf person on a verbal ability intelligence scale could imply that the individual has limited verbal aptitude, or it could merely reflect the fact the deaf person has been denied the opportunity to acquire verbal and social knowledge due to the person's hearing loss. (p. 76)

To try to accommodate for the seeming lack of verbal ability the D/HH demonstrated on verbal intelligence tests, separate "deaf" norms were established for some tests to compensate for the discrepancy in scores between the D/HH and the hearing samples. Deaf norms were available on five intelligence tests including the Wechsler Intelligence Scale for Children-Revised, the Hiskey-Nebraska Test of Learning Aptitude, the Sniijders-Ooman Nonverbal Intelligence Test, the Pinter Nonlanguage Test and the Raven's Progressive Matrices. Though the use of deaf norms created a slight difference in the *distribution* of IQ scores between the D/HH and hearing samples, the use of deaf norms did not actually yield IQ scores that were significantly different than the use of norms based on hearing individuals (Braden, 1992, 1994, 2005). Additionally, samples of D/HH students used in creating deaf norms were generally all residential students without

a randomly selected or stratified sample. Further, the average age of the D/HH students was significantly higher than that of the norms developed with hearing samples. All these factors created unequal comparisons between the two normative groups (Braden, 1994).

Thus, the use of traditional intelligence tests with the D/HH has been highly controversial. As Braden (1994) emphasized, in assessing the cognitive ability of D/HH, traditional intelligence tests with high verbal content, threaten the utility and validity of testing scores for this unique population. This issue has been debated and continues to be debated.

There are several significant reasons for questioning the validity of using intelligence tests with the D/HH. The first reason is that the D/HH is a very low incidence population which creates several complicating variables, namely that there are a limited number of individuals on which to conduct the research necessary to norm, standardize and validate a measure, and also that it is not as profitable for a publishing company to create a test specifically for such a small population.

A second reason is that there are many subgroups within the D/HH population, in terms of level of hearing loss, parental hearing status/loss, age of onset of hearing loss, etiology of hearing loss, the frequency, duration, and intensity of language exposure and the presence or absence of additional disabilities (Braden, 1994; Cone-Wesson, 2003; Marschark, 2003). Similarly, there are several communication modalities used by the population including auditory/spoken language and lip reading, ASL, Signed Exact English (SEE), and Cued Speech, etc., further decreasing the possible sample sizes of any investigational groupings. A third reason is that very few researchers possess both the knowledge base and skills necessary to work with the D/HH population and also the

knowledge base and skills necessary to investigate the statistical properties of intelligence measures (Maller, 2003).

Despite all of the differences within the D/HH population and in light of initial finding that verbally loaded intelligence tests tend to produce lower scores in the D/HH than those of hearing samples, researchers and educators have continued to search for a method, process, or test by which the intelligence of the D/HH can be validly measured. Again, such a method is desired because of a need to identify abilities at both of the spectrum: giftedness and intellectually disability. This type of information is also helpful to inform educational intervention, assist with employment programming, and monitor individual progress.

Overview of Assessment Research

In a 1992 review of research regarding the cognitive assessment of individuals identified as D/HH, Jeffery Braden analyzed several decades of published, unpublished, and pre-published works to determine which intellectual assessments had been used, what the outcomes of the assessments were, and if there were any relationships between the assessments and the outcomes. In 1994 Braden re-published this review as a book chapter in *Deafness, Deprivation and IQ*, examining not only cognitive functioning, but also academic achievement, adaptive behavior, and the implications of hearing loss on development.

Because Braden's 1992 review of the existing literature was so extensive, it is considered the stand-alone seminal history of deafness and cognitive functioning prior to 1992. His work included several key findings, beginning with the notion that research in the field of intellectual assessment for the D/HH was growing at a slow pace and was

often conducted--and then the results shared--in an arena isolated from mainstream psychology. Braden made the following observations:

In the same way deaf and hard-of-hearing people are isolated from the mainstream of society, research in this area is isolated from the mainstream of psychology. . . . This may be due to the poor quality of the research, but it also could be due to resistance among psychological journals to allocate space to 'low incidence' topics such as deafness and hearing disorders. The net effects of the journalistic isolation are that psychologists who do not read journals related to deafness and hearing disorders are unlikely to be familiar with the research, and the field of deafness is unlikely to attract high-quality psychological researchers because of the low visibility of research about deafness and hearing disorders. (p. 89-90)

Braden's 1992 review included 285 separate studies which contained 234 separate samples of D/HH individuals. An overall look at the research studies demonstrated that older studies reported lower mean IQ's than did more recent studies. This finding might lead one to believe that, in respect to cognitive ability over time, individuals identified as D/HH are *catching up*, becoming similar to their normal-hearing peers. However, perhaps a more reasonable assumption is that the quality of studies is improving over time and similarly, the ability to accurately evaluate the intelligence of D/HH. Furthermore, regardless of the study being quantitative or qualitative in nature, less rigorous studies (poor quality) were associated with lower IQ scores, regardless of how recently the study was conducted.

A summary of the 324 reports of mean IQ were analyzed. There was a discrepancy between the findings in 246 quantitative studies when compared to the

summary of the 52 qualitative studies. The quantitative reports provided a higher estimate of D/HH individual's intelligence than did the summary of the qualitative reports.

Further, the summary of results from the qualitative reports showed that the distribution's mean of intelligence in the D/HH population, was somewhat below the distribution's mean for the hearing population, which supported the cognitive inferiority theory of the D/HH, previously discussed.

Wechsler Performance Scales. The most popular assessment instrument subtests, which each contained highly verbal content and directions, were the Wechsler Performance Scales (Braden 1992, p. 85). And, as suggested earlier, when assessing D/HH, tests with higher verbal content yield lower IQ's than do tests with low verbal content. Even though verbal tests yield intelligence scores that are about one standard deviation lower for the D/HH than nonverbal scores, the research suggests that standardized tests may even overestimate the verbal reasoning skills of the D/HH. Thus the average discrepancy between verbal reasoning and nonverbal reasoning abilities of the D/HH could be greater than the standard deviation difference suggested by Braden's review.

Braden's summary also indicated that differing communication methods for administering tests appear significantly related to measured intelligence. The net effect of administration procedures appears to be that signed administration methods yield higher IQ's than pantomimed, oral, or written methods. Though this finding only applies when the tests were administered to D/HH individuals who used sign language as their primary mode of communication.

Braden identified a trend, showing that since the inception of intelligence testing with D/HH children, studies reported that D/HH children in residential programs scored

lower on nonverbal IQ tests than children in nonresidential programs. (Residential program are schools in which the students not only attend each day, but stay in dormitories on the school property each night during the school week and then generally return to their homes on the weekends.)

Many authors assumed that assessment methods producing higher IQs were better than methods producing lower IQs. So, as earlier indicated, many educators have suggested using special norms when administering cognitive tests to D/HH students. However, this recommendation has not been supported with empirical data showing greater accuracy or better analysis of intellectual abilities. Overall, greater accuracy has not been achieved when special norms are used (Braden, 1992).

Non-verbal tests. Braden (1992) indicated that though procedural factors were considered in this summary the influence of participant characteristics (age, degree of hearing loss, site from which participants were sampled) was not. Thus, further research needs to be conducted. One such analysis would be the degree to which participant traits are associated with specific assessments and assessment practices, as in the seemingly common practice to administer the Leiter Nonverbal Intelligence test (Leiter) to low-functioning children and Wechsler scales to the higher functioning, which may skew results. Another is a study of whether motor-free nonverbal tests are interchangeable with performance tests for assessing the intelligence of D/HH people.

Key Research Findings Regarding IQ Assessment of D/HH

Even with the backing of research findings, information and recommendations related to the cognitive assessment of D/HH students must be interpreted with caution. Braden (1992) stated, “Because of the dearth of studies comparing multiple intelligence tests within the same sample, it is not known whether the differences among tests are a

function of sample selection, test characteristics (e.g. subtest complexity, norms), or administrative procedures” (p. 91).

Building upon the studies reviewed by Braden (1992), other authors have attempted to examine the relationship between D/HH students and their hearing counterparts. Additionally, researchers have also made comparisons between groups of D/HH students on various cognitive measures (Hooper, 2002; Hughes, Sapp, & Kohler, 2006; Hunt, 1997; Krivitski, 2000; Mackinson, 1996). The most common assessments used in such studies were the Universal Nonverbal Intelligence Test (UNIT) and the Wechsler Intelligence Scales for Children (WISC).

Validity of IQ Testing with D/HH

Considering that the WISC and UNIT are the two most widely studied cognitive assessment instruments with D/HH students, the attributes of these instruments must be scrutinized. In particular, Braden and Athanasiou (2005) reviewed *language reduced* measures of intelligence, noting the importance of considering two important issues: does the test measure what it purports to measure and are the results of the assessment consistent across time. These two issues, validity and reliability, are vital attributes of assessment. Likewise, these two issues must be considered when selecting assessment instruments for measuring the cognitive ability of D/HH students. This section investigates research related to cognitive assessment of D/HH populations, reviewing aspects of test validity.

Maller and Braden (1993) established that the WISC-III had both construct and criterion-related validity for 30 D/HH junior high and high school students. All of the students used sign language to access the educational curriculum. The students were administered the WISC-III and the Stanford Achievement Test, Hearing-Impaired Edition

(SAT-HI; Allen, 1986). The mean Full Scale IQ standard score was 92.17, and the mean Verbal Scale standard score was 81.12 for the D/HH sample, both of which were statistically significantly different than that of the WISC-III standardization sample. The mean Performance Scale standard score for the D/HH sample was 105.83, which was not statistically significantly different than that of the standardization sample. Maller and Braden reported that these results were similar to results found with the WISC-R.

Maller and Braden (1993) also found moderate correlations between the Performance Scale scores and the Stanford Achievement Test-Hearing Impaired version (SAT-HI) scores, while the Verbal Scale scores correlated highly with the SAT-HI scores, and all the correlations were statistically significant. Additionally, they concluded that even though this was a small sample of D/HH adolescents, the results were useful because they replicated previous findings. More specifically, similar to studies with the WISC-R, the scores from the WISC-III Performance Scale could be used as a measure of cognitive ability for D/HH adolescents and the Verbal Scale scores could be used to measure academic abilities.

One study included in the Braden (1992) review described earlier was published in 1984 by Miller. This study indicated that D/HH students had normal Verbal IQ scores (mean = 96) on the WISC-R. Miller's finding was interesting because the results were significantly different from previous studies. However, even with Miller's results added into the mix of all the other studies, the mean Verbal IQ score for D/HH individuals was at least one standard deviation below that of hearing peers.

Then, in 1994, Chovan and Benfield replicated the method in Miller's 1984 study. The researchers randomly assigned 27 D/HH students to three groups that were each then administered the WISC-R Verbal Scale using one of three different communication

methods (American Sign Language and Pidgin Signed English, Signed English, and Maximum Communication System). Chovan and Benfield found that none of the three groups reached a mean Verbal IQ of 96. In fact, none of three groups even came close as they scored 74, 76, and 80 respectively, validating the results that one would expect after having read the Braden (1992) review.

Cash (1994) investigated the concurrent validity of the WISC-R and the Test of Nonverbal Intelligence-Second Edition (TONI-2), as well as the predictive validity of the WISC-R scores on the Stanford Achievement Test 8th edition (SAT-8), with a sample of D/HH students. When comparing the two cognitive measures, there were no significant discrepancies on scores for this sample of D/HH students. Further findings indicated a lack of significant difference between D/HH students (98.043) and normative scores from the TONI-2. Furthermore, the Performance Scale of the WISC-R was found to be a moderate predictor of SAT-8 scores, with correlations ranging from .36 on the Total Language scale and .53 on the Total Math scale.

This study seemed to indicate the possible use of nonverbal intelligence as a predictor of academic success for D/HH children, and the TONI-2 as a measure which produced scores that were predictive of academic achievement in D/HH students. This study indicated limitations with sample size and the sampling methods used, as D/HH students who were included may have been tested because they were suspected of having additional disabilities.

In a very similar study, Slate and Fawcett, (1995) reviewed file data concerning the relationship of the WISC-R Performance Scale, WISC III Performance Scale, and the WRAT-R. The test scores of 47 D/HH students were compared, and since all but three received ongoing three-year psycho-educational evaluations, the necessary test scores for

this analysis were found the students' files. Twenty five students communicated using spoken language, while 22 used Total Communication (TC), meaning to use both spoken English and sign language to communicate.

It was found that the mean WISC-R (91.0) and WISC-III (88.0) scores were more than one-half of a standard deviation below the mean for the standardization samples, but were highly correlated with one another ($r = .93$). The students in the sample who communicated using TC achieved higher Performance IQs than their counterparts who communicated orally, but the difference was not statistically significant. WISC-III Performance IQ had some relevance to the WRAT-R subscales, especially in math ($r = 0.64$). This finding supported previous findings for the predictive validity of the WISC-R and the WRAT-R. Limitations to the study included the small sample size and the restricted geographical size.

Mackinson (1996) examined the concurrent validity of the Test of Nonverbal Intelligence – Second Edition (TONI-2) and the WISC-III Performance Scale. Twenty-seven D/HH participants were administered the TONI-2 and the required subtests on the WISC-III with 24 of the participants being given the optional subtest, Symbol Search. The WISC-III Performance Scale and the TONI-2 were administered by a trained clinician with interpreter-level (or near interpreter level skills).

Pearson-Product Moment Correlation Coefficients for the WISC-III Perceptual Organization Index score and Processing Speed Index scores were .67 and .57, respectively. Both demonstrated moderate correlations. Mackinson therefore noted, “These findings suggest the WISC-III PIQ is a good predictor of the TONI-2 score and that these measures may assess similar constructs, though one is a *nonverbal-administered, motor-free* measure and the other is a *motor-intensive* one” (p. 83).

A 1997 study (England) examined the predictive and concurrent validity between the Universal Nonverbal Intelligence Test (UNIT) and the Test of Nonverbal Intelligence – 2nd Edition (TONI-II) and achievement, comparing them with the Peabody Individual Achievement Test-Revised (PIAT-R) and the Stanford Achievement Test-Eighth Edition (SAT-8). Between the TONI-II Total Raw Score and the UNIT Memory and Reasoning Composite scores, the Memory Composite of the UNIT was the best predictor of reading comprehension in the PIAT-R and math achievement on the SAT. The Symbolic Quotient of the UNIT was a better predictor than the Nonsymbolic Quotient of overall achievement when compared to the SAT and the PIAT-R. With the exception of PIAT-R Reading Comprehension, the TONI-II was not a good predictor of achievement.

Krivitski (2000) found that D/HH and hearing children performed similarly on the Universal Nonverbal Intelligence Test and its subtests. Thirty-six D/HH children and 39 hearing children were administered all six subtests of the UNIT. The Object Memory subtest had the lowest mean score for both groups, while the Cube Design subtest had the highest mean for both groups. Using discriminant analysis, Krivitzki concluded that Cube Design was the subtest that provided the most information about the cognitive performance of both D/HH and hearing children. For both groups, the Symbolic Quotient had the lowest mean, while the Nonsymbolic Quotient had the highest mean.

Krivitski concluded that as the UNIT assesses a wider variety of skills than most performance or nonverbal measures, it provides more information about a child's skills and areas of needs, making the UNIT a good tool to use with D/HH children. However, a limitation of the study was that the UNIT was not compared with other measures and that demographic factors of the D/HH sample, especially, were not considered since age of

onset of hearing loss, degree of hearing loss, parental hearing status as well as other factors could strengthen the finding's generalizability.

Maller and French (2004) compared a D/HH population sample to the standardization samples of the Universal Nonverbal Intelligence Test. They focused on two theoretical models: a primary two-factor (Memory and Reasoning) model and a secondary two-factor (Symbolic and Non-symbolic) model.

The 102 D/HH participants, ages 5 through 17 were students from four sites (located in the southeastern, western, and mid-western United States). The D/HH sample included 44 males and 58 females who were identified as members of the following racial or ethnic groups 74 White, 15 African American, 7 Hispanic, 3 Asian/Pacific Islander, and 3 "other." (p. 650)

In spite of the fact that research on D/HH examinees suggests that they have performance IQ's which are consistent with their hearing peers, they may have problems with motor-free nonverbal intelligence tests. For instance, on the Analogic Reasoning subtest, the D/HH scored lower than the standardization sample.

Previous research confirms that the D/HH "have difficulty with short-term memory tasks and obtain lower scores on nonverbal, visual short-term memory tests, possibly due to interference and time required for encoding verbally mediated tasks by signers or, similarly, the lack of speech-related encoding strategies" (Maller & French, 2004, p. 656). They speculate it is possible that the finding based on their D/HH sample "was not necessarily generalizable" (p. 657) and that some members of their sample may have had additional unidentified disabilities. Moreover, there is some concern that the UNIT gestures, which are unlike sign language, used to demonstrate the subtest tasks may be unclear or ambiguous to D/HH children.

A recent study which is very similar to the proposed study was a comparative examination of scores on the UNIT and the WISC-III Performance Scale in a sample of 32 D/HH students which were residents of a state school (Hughes, Sapp, & Kohler, 2006). Though this study was neither published, nor peer-reviewed because of its similarity to the proposed study, it is important to include. A significant finding was that the mean scores on both tests were approximately one standard deviation below the mean of the normative samples of both tests.

The results indicated that there was not a significant difference between the students' mean Full Scale score on the UNIT (87.44) and the Performance Scale of the WISC-III (84.28). However, when the 20 middle school students' scores were separated from the 12 elementary school students' scores, there was a difference in mean scores, as the middle school students had significantly higher scores on the UNIT than the WISC-III. Significant correlations among the subtests of both measures demonstrated "conceptual overlap" (Hughes, Sapp, & Kohler, 2006, p. 9).

The information previously presented serves as a summary for the decade of validity studies. These studies demonstrate that there is a significant difference in the scores of D/HH individuals on verbal versus performance-type intelligence tests. Most researchers have concluded that for D/HH individuals, there is more evidence of the validity of the use of scores from performance rather than from verbal measures of cognitive functioning. However, since verbal scores appeared to be better at predicting academic achievement in the D/HH there was still value in administering the verbal scales.

Other findings included that there have been several nonverbal intelligence tests shown to be valid measures of intelligence with the D/HH, including the TONI and the

UNIT. But, when those two tests are compared to one another, the UNIT is a better predictor of the academic achievement of the D/HH (England, 1997). However, as Maller (2003) indicated, the results are far from conclusive; some studies show the scores of D/HH individuals to be similar to the normative samples, while other studies have shown them scores to be lower.

Factor Analysis of IQ Assessment Scores of D/HH

Thurstone (1931), in an effort to summarize data and make sense of large data files, introduced *factor analysis*. Factor analytic techniques categorize and reduce information, helping make sense of data, providing a context and relationship among data.

Sullivan and Montoya (1997) factor analyzed the WISC-III on a sample of 106 D/HH children ages six through sixteen. Two major factors emerged: Language Comprehension and Visual-Spatial Organization. There were no differences in Verbal, Performance or Full-Scale IQs between children in mainstream vs. residential placements; signing test administrator, use of sign language interpreter or oral only directions; children whose communication mode was oral, ASL or signed English; boys and girls; or children with moderate, severe or profound hearing losses. Children with known etiologies (i.e., meningitis, perinatal complications, rubella, cytomegaloviral inclusion, or genetic anomalies) earned significantly lower Performance IQs and Object Assembly scores than children with unknown etiologies. This is consistent with previous research that demonstrated that children with high-risk etiologies for neurological dysfunctions earn lower IQ scores than children without these etiologies (Braden, 1994; Karchmer & Mitchell, 2003).

The mean IQs for the entire sample were Verbal, 75.35 ($SD = 17.55$); Performance, 100.63 ($SD = 19.48$); and Full Scale, 86.22 ($SD = 17.37$). The mean verbal IQ was 1.67 standard deviations lower than that of the standardization sample, while the Performance IQ was equal to that of the standardization sample.

Sullivan and Montoya (1997) suggested that the “historic taboo” (p. 320) against using verbal intelligence tests with D/HH children be re-thought because: (a) the majority of them are educated with hearing peers in settings where they must compete with hearing students and academic subjects are language based; (b) verbal IQ is a better predictor of reading and math achievement among D/HH than performance IQ; and (c) improved literacy, numeracy, and face-to-face communication skills will be related to better paying jobs in adulthood.

In another study published the same year, Maller and Ferron (1997) analyzed the construct validity of the WISC-III. The researchers compared the four factor structure found for the standardization sample (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility and Processing Speed) with that of a sample of children with severe to profound hearing losses. One hundred and ten students from either a residential program, a self-contained program housed in a regular education facility or a residential day program, all of whom used sign language, were included in the study. Maller and Ferron concluded that that four factor model fit for standardization sample and the D/HH sample. However, Maller and Ferron noted,

Some subtests did not indicate the latent construct of intelligence in a similar manner across samples, with relationships between factors also affected by group membership. . . . The administration of some subtests via ASL may interfere with

their intended constructs. For example, the Digit Span subtest becomes a visual, rather than auditory, short-term memory task. (p. 990)

These two studies found dissimilar results, though both factor analyzed the WISC-III. The first found a two factor model fit the sample of D/HH children best, and concluded that since verbal scores correlate with academic achievement in D/HH individuals, they should still be administered. The second study found that the four factor model of the WISC-III fit the sample of the D/HH as well as it did the standardization sample, but that there were differences in the way some of the subtests described the two samples.

Item Analysis of D/HH Students' Responses on the WISC and UNIT

Maller's three studies (1996, 1997, 2000), described below, investigated differences in the way individual items function for D/HH children compared to hearing children on the WISC-III and UNIT. Though differences were detected on several WISC-III items, differences in the way UNIT test items functioned for the two groups were not detected, even when younger hearing children were compared with older D/HH children.

In an effort to determine if D/HH children score lower on verbal intelligence measures because of a delay in language development, Maller (1996) compared D/HH children who use sign language and a sample of younger hearing children on individual items within the Verbal subtests of the WISC-III. Her hypothesis was that if language delay was the reason for the poor performance, the D/HH children would perform similarly to children at a lower language level. There were 110 D/HH children in the sample ranging in age from eight to 16 compared with the standardization sample for the WISC-III ages 6 through 14 years-old.

The items were translated into American Sign Language (ASL) or Pidgin Signed English (PSE) and then back translated into English to ensure the D/HH were receiving the same administration as the standardized version. (Most of the Vocabulary subtest items had to be fingerspelled as there was no direct ASL interpretation, so those items were identical to the standardized version.) “Progressively younger (i.e., by one month younger at a time) subsamples of hearing children were analyzed until the subtest raw score and Rasch logit ability mean and standard deviation estimates were equal to those of the deaf sample” (Maller, 1996, p. 156).

On all five of the subtests, the D/HH children were older than the hearing children with the same ability level, for example, the D/HH children’s mean scaled score on the Vocabulary subtest was two standard deviations below the mean of the hearing children. However, because there were many items that did not fit the D/HH sample, but did fit the younger hearing sample, there is evidence that the two groups differ in response pattern as well. This meant that the items did not retain their degree of difficulty for the two samples, as some items harder for the younger hearing group were easier for the D/HH sample and vice versa. Maller concluded that, “. . . it is a questionable practice to test D/HH children using Verbal scale items calibrated on a sample of hearing children, and then to report subtest scale scores that equal those expected of much younger children” (Maller, 1996, p. 163).

Maller (1997) matched hearing and D/HH samples of children by age and performance IQs and then attempted to detect differential item functioning (DIF) within five subtests from the Wechsler Intelligence Scale for Children-Third Edition (WISC-III). While this study compared mostly verbal subtests, one performance subtest, Picture Completion, was included.

A “. . . one-parameter Item Response Theory, or Rasch model was used to analyze item level data. . .” (p. 303) as the “standardized difference in difficulty and anchor item...” (p. 299) determined DIF. Table 1 demonstrates the number of items that exhibited DIF for D/HH children on each subtest examined in this study.

The study indicated that the subtests “appeared to differentially measure intelligence for deaf children” (p. 310). However, the study also indicates that there did not appear to be a good fit between subtest items for children who were D/HH, indicating that overall ability was not a good predictor of passing or failing an item based on its difficulty, even while ability was a good predictor of passing or failing an item based on its difficulty for hearing children. While the significance of differences in favor and against D/HH students appears to cancel each other out in the overall subtest raw score, Maller (1997) stated, “ceiling rules may be inappropriate because individuals may not necessarily reach the items that exhibit DIF in their favor” (p. 311).

Maller concluded that her study indicated a need for D/HH norms to be established following a revision of subtests from the WISC-III removing items that are biased for and against D/HH students. She also concluded that psychologists and educators need to adjust attitudes about using verbal intelligence tests with D/HH students and identify an alternate method of assessment.

Following the release of the Universal Nonverbal Intelligence Test (UNIT), Maller (2000) researched the individual items from four of the UNIT subtests, including Symbolic Memory, Spatial Memory, Analogic Reasoning and Object Memory, to determine if there was differential item functioning (DIF) for children who were profoundly D/HH or hearing. “Items were screened for DIF via the Mantel-Haenszel (MH) DIF detection method” (p. 224).

Table 1

D/HH Students' Number of WISC-III Subtest Items Exhibiting Differential Item Functioning

Subtest	Picture Completion	Information	Similarities	Vocabulary	Comprehension
DIF in favor of D/HH	2	3	4	6	3
DIF against D/HH	4	3	3	8	3
Poor anchored fit statistics	2	10	0	3	0

Note. This table is based on Susan Maller's 1997 study.

This study found that no items on the UNIT exhibited DIF for either the D/HH or hearing children. Maller (2000) concluded that, "Psychologists now have a tool for assessing intelligence in D/HH children, with some evidence that the items in the subtests studied are invariant for deaf children" (p. 250) The study indicated a need for further research of the psychometric properties of the UNIT.

Subcategories of D/HH: Implications for Interpreting IQ Scores

Though deafness is a low incidence disability, D/HH individuals comprise a very heterogeneous population, with many subcategories. D/HH individuals have varying degree of hearing loss from mild to moderate, to severe and profound. Students attend educational programs either based solely on American Sign Language, or on using residual auditory abilities to develop students' oral skills, or a combination of the two programs. Additionally, D/HH students are educated in self-contained classes by teachers

endorsed to teach D/HH students; in regular neighborhood schools, with or without an interpreter, depending on the student's level of communication needs; and at residential schools where students spend the entire week at the school, only going home on weekends with their families.

Understanding if and/or how these subcategories of D/HH students' performance on cognitive measures differ has been investigated in the research. Hunt (1997), dividing the D/HH students into three groups based on type of communication, considered the impact of communication on D/HH students' cognitive test scores. Braden, Maller, and Paquin (1993) examined differences in cognitive test scores, categorizing scores based on the type of educational program attended by D/HH students.

Form of Communication Utilized by Student

In Hunt's 1997 study, results from the WISC-III for D/HH children using one of three forms of communication (Cued Speech, Total Communication, and Oral/Aural) were evaluated (Hunt). The focus of the study was to determine the ". . . appropriateness of considering the Verbal Intelligence Quotient (VIQ) of the WISC-III as a valid measure for deaf children" (p. 1). This study consisted of a limited sample size of 51 children including 17 children, in each of the three categories of communication method.

Based on the results, 82.4% of the total children had a 15 point or greater discrepancy between the PIQ of the WISC-III and the VIQ of the WISC-III. No significant differences were found between the three communication groups for either the VIQ or the PIQ. While the sample size was small, the results of this study did support previous findings. Hunt concluded that the findings ". . . suggest that the Verbal scale of the WISC-III is not a valid measure of the [D/HH] child's true verbal abilities as defined by Wechsler (1944)" (Hunt, 1997, p. 56).

Type of Educational Program

To further examine the effects of education placement on IQ scores, especially residential versus day school, Braden, Maller and Paquin (1993) examined multiple scores, over time, for each of 208 students. Sixty-two of the participants attended a regional day program, meaning they attended classes for the D/HH within regular schools. One hundred and five students attended a residential school for the D/HH at which they lived Monday through Friday returning to their homes on weekends, and 41 attended the day program at the same residential school. The Wechsler scales were used as the measurement tools.

The results showed significantly increased Performance IQ standard scores over three year's time for the residential students living on campus and for the students attending classes for the D/HH located in regular schools. The day students at the residential school showed no change in PIQ scores over time. These results refuted the trend found in Braden's review that residential schools have a debilitating effect on D/HH children.

Conclusions and Future Directions

As studies of cognitive ability in the D/HH became more sophisticated, the belief that D/HH individuals were cognitively inferior was replaced by an understanding that, in comparison to the general population, there are fundamental differences in the way D/HH children perform on intelligence tests. After data made it clear that verbally loaded tests produced scores that had limited evidence of validity as measures of the intellectual ability of the D/HH but could predict academic achievement, researchers attempted to find cognitive assessments that were more appropriate for use with D/HH individuals.

Research studies began separating, first, the specific categories of the D/HH communities (such as residential vs. day school, ASL vs. Auditory/Oral language users) for examination and comparison; second, they began breaking down the nonverbal cognitive tests into categories such as test requiring motor abilities and tests that were motor free, nonverbal tests requiring reasoning versus nonverbal tests requiring short term memory; and finally, they examined the individual items for comparison.

Multiple studies have investigated the validity of the Wechsler scales with the D/HH and recently several studies have examined the validity of the use of the UNIT with D/HH population. However, no study to date has been published that has compared the WISC-IV and the UNIT.

Thus, the current study is a logical extension of both recent research trends and the subsequent questions posed by those conducting the research. And, it will provide the data necessary to make a determination about the appropriateness of using the UNIT without and instead of the WISC-IV with the D/HH students at USDB as an eligibility criterion for both special education and state social services. Finally, the current study directly addresses Braden's 1992 recommendation that more research is needed which compares multiple intelligence tests within the same sample, by comparing the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV) with the Universal Intelligence Test for Children (UNIT).

Research Questions

1. What is the nature of the distribution of WISC-IV PRI scores in a robust, convenient sample of USDB D/HH students?
2. What is the nature of the distribution of UNIT Standard Battery Composite scores in a robust convenient sample of USDB D/HH students?

3. Is there a significant difference between a robust convenient sample of USDB D/HH students and the WISC-IV standardization sample in terms of the WISC-IV PRI score?
4. Is there a significant difference between a robust convenient sample of USDB D/HH students and the UNIT standardization sample in terms of the UNIT Standard Battery Composite score?
5. What is the correlation between the WISC-IV PRI score and the UNIT Standard Battery Composite score in a robust convenient sample of USDB D/HH students?

Methods

Selection of Participants

The proposed research study was archival, as data were retrieved from existing files of D/HH students enrolled at the Utah Schools for the Deaf and the Blind (USDB) between 2005 and 2007. USDB serves students across the state of Utah; however, not all D/HH students attend or receive services from USDB. In Utah, more than 75% of D/HH students are mainstreamed in their neighborhood schools, but most of those students receive outreach services from USDB. The majority of students who attend USDB require more intensive service than their neighborhood schools or an outreach teacher can reasonably provide. USDB students are often identified with significant language delays.

For this particular study, USDB students' test scores were selected for data analysis if the student met three criteria. The first criteria was that the student was six years-of-age or older, but less than 17-years of age. This age range was selected to align with the overlapping age range of the UNIT and WISC-IV normative samples. (The UNIT is for children five years zero months to 17 years 11 months and the WISC-IV is for children six years zero months to 16 years and 11 months.) USDB students' data were therefore selected based on age, but not selected or sorted based on gender, ethnicity, socioeconomic status or secondary special education classification of the student.

The second criteria was that the students were administered both the Standard Battery of the UNIT and the Perceptual Reasoning Index subtests from the WISC-IV by USDB school psychologists, within a routine triennial psychoeducational evaluation for the purpose of special education classification or reclassification.

The third and final criteria for inclusion was that participating students were required to have a current consent form on file. Only approximately 1/3 of the D/HH students enrolled in USDB, met all three criteria, making this a convenient sample.

Sample Demographics

The 61 participants whose data were used for this study were D/HH students at USDB who were administered the WISC-IV and the UNIT during the past three years for routine special education eligibility evaluations. The students ranged in age from six years zero months through 16 years 11 months with a mean age of 10.90. Thirty-five of the students were in elementary school and 26 were in junior high and high school. In regard to gender, 34 were males and 27 females. Parents of six students spoke Spanish; the remaining 55 students' parents were English speaking. As their preferred method of communication, 29 of the students were oral, 22 used total communication and 10 used American Sign Language. Table 2 describes the participant demographics.

Setting

Testing was administered in conference rooms and testing rooms within the various schools attended by participating students, or in the USDB offices of the examiners (school psychologists).

Instruments

UNIT

The UNIT “is a language-free test that requires no receptive or expressive language from the examiner or the examinee” (McCallum & Bracken, 1997, p. 268). The examiner uses gestures (e.g., pointing, nodding, and shaking the head) to communicate with the examinee. The UNIT also provides several demonstration items for each subtest so that the examiner can show the examinee how to accomplish the task. The UNIT was

designed to provide a fair and valid measure of intelligence for children ages five years zero months to 17 years 11 months who would be disadvantaged by traditionally language-loaded tests. With the intent to design a culture-free instrument, the test's authors chose universal gestures to provide instructions to examinees and to demonstrate subtest activities. Examinees' responses are also nonverbal.

Table 2

Participant Demographics

Demographic		<i>N</i>	Percentage
Gender	Male	34	55.7
	Female	27	44.3
ELL	Not ELL	55	90.2
	ELL	6	9.8
Program	TC	22	36.1
	Oral	29	47.5
	ASL	10	16.4

Note. *N* = 61.

The UNIT Standard Battery includes four subtests. The Symbolic Memory subtests show the child increasingly complicated groups of stick figure person cards, and then must remember the order of the figure cards after they are taken away. The Cube Design subtest requires that the child look at increasingly complicated two dimensional

pictures of blocks arranged into a geographic design, and then recreate it with real blocks. The Spatial Memory subtest requires the child to look at increasingly difficult arrangements of dots on spaces within a matrix. The matrix is then shown to the child without any dots on it and the child must supply the dots in the correct spaces. The Analogic Reasoning subtest requires the child to choose the correct solution from several choices to increasingly complex matrices that are missing one small piece. The scores for each subtest have a mean of ten and a standard deviation of three.

UNIT standardization sample. The UNIT standardization sample included 2,100 children, ranging in age from five years zero months to 17 years 11 months 30 days (Bracken & McCallum, 1998, p. 19). The 1995 census data was used to develop a representative sample stratified by the following characteristics: gender, race, Hispanic origin, region, community setting, educational placement, special education services and parental education variables. Based on information provided by Bracken and McCallum, the participants came from 38 states and 108 sites (p. 20). Further, the standardization sample included a subsample of students receiving special education services in which 0.2% of the students were classified as D/HH. This percentage of D/HH students was consistent with census data, as well as educational data from the U.S. Department of Education.

UNIT standard battery. The UNIT Standard Battery consists of four subtests, each measuring a somewhat different aspect of nonverbal intelligence: (a) Symbolic Memory is primarily a measure of short-term visual memory and complex sequential memory for meaningful material; (b) Cube Design is primarily a measure of visual-spatial reasoning; (c) Analogic Reasoning is a measure of symbolic reasoning and, (d)

Spatial Memory is primarily a measure of symbolic reasoning. The subtests have a mean of ten and a standard deviation of three.

The student's performance on these various measures is summarized in five scale scores: (a) the Memory Quotient, which includes Symbolic Memory and Spatial Memory; (b) the Reasoning Quotient which includes Cube Design and Analogic Reasoning; (c) the Symbolic Quotient which includes Symbolic Memory and Analogic Reasoning; (d) the Nonsymbolic Quotient which includes Cube Design and Spatial Memory; and then the Standard Battery Full-Scale IQ (FSIQ) which includes all four subtests. The scale scores have a mean of 100 and a standard deviation of 15 (Bracken & McCallum, 1998).

Reliability of UNIT. Several methods were used for estimating the reliability of scores obtained from the UNIT. The first is Internal Consistency, on which, the median of the average subtest reliability across ages is .80 for the Extended Battery which consists of all six subtests. The Full Scale score of the UNIT, which is the estimate of general intelligence, has a reliability coefficient of .93 for the Extended Battery. The reliability coefficients for each of the four subtests of the Standard Battery are as follows: Symbolic Memory = .85; Cube Design = .91; Spatial Memory = .92; Analogic Reasoning = .79; Object Memory = .76; and Mazes = .64. (Bracken & McCallum, 1998, p. 100). Table 3 describes the reliability coefficients for the UNIT.

To obtain test-retest stability, 197 children between the ages of 5 and 17 were administered the UNIT and then about three weeks later took it again. The mean interval between tests was 20.3 days, and there was an average of 4.8 points of practice effect on the Extended Battery, with a corrected correlation coefficient of .85 (Bracken & McCallum, 1998, p. 108). Overall, the authors of the UNIT concluded that the reliability

coefficients indicate, “a high level of measurement precision in terms of internal consistency and stability over time” (Bracken & McCallum, p. 110).

Table 3

UNIT Reliability Coefficients

Composite/ Subtest	Extended Battery	Symbolic Memory	Cube Design	Spatial Memory	Analogic Reasoning	Object Memory	Mazes
Number of Items	n/a	30	15	27	31	30	13
UNIT Reliability Coefficient	.93	.85	.91	.92	.79	.76	.64

Validity of UNIT. Internal and External Validity were also estimated for the UNIT. Internal validity was measured by intercorrelations. The intercorrelations between the Full Scale score and the subtest scores ranged from .76 on the Spatial Memory subtest and .70 on the Object Memory subtest, except for the Mazes subtest which was .52 (Bracken & McCallum, 1998, p. 125). Factor analyses revealed that two major factors emerged, that of *memory* and *reasoning*, but that the “one-factor *g* model appears to provide the most parsimonious fit to the data” (Bracken & McCallum, p. 131).

The UNIT’s external validity was estimated by comparing UNIT test scores with other measures of cognitive functioning. The Full Scale UNIT score correlated strongly with the Wechsler Intelligence Scale for Children – Third Edition (WISC-III) with a corrected correlation coefficient of .83 for a sample of children with identified learning

disabilities; .88 for a sample of children with mental retardation; .88 for a sample of intellectually gifted children; and .65 for a sample of Native American children (Bracken & McCallum, p. 136). Table 4 describes the UNIT's external validity coefficients.

Table 4

UNIT External Validity Coefficients

	Learning Disability	Mental Retardation	Intellectually Gifted	Native American
Number of participants	61	59	43	34
WISC-III and UNIT correlation	.83	.88	.88	.65

When the two tests, the UNIT and the WISC-III were compared with the Woodcock Johnson Tests of Achievement – Revised (WJ-R) had high correlations with the Broad Mathematics, Broad Reading and Broad Written Language Cluster scores. When compared with the Broad Cognitive Ability Composite score of the Woodcock Johnson – Revised Tests of Cognitive Ability, the UNIT a corrected coefficient of .82 was obtained. (Bracken & McCallum, 1998, pp. 122–152).

A sample of 106 D/HH children (60 females and 46 male) who were receiving special education were administered the UNIT, and their scores were compared to those of the standardization sample. All of the students in the sample attended a school for the D/HH which required that the students had delayed language, lower academic performance, and an inability to communicate effectively when compared with age-mate students who were not D/HH. The sample was matched by age, sex, race, ethnicity and

parent education to the standardization sample. The mean Full Scale score difference was 8.01 points higher for the standardization sample than the D/HH sample on the Extended Battery (Bracken & McCallum, 1998, p. 192).

WISC-IV

The WISC-IV is an individually administered clinical instrument for assessing the intellectual ability of children ages six years zero months through 16 years 11 month 30 days. The WISC-IV consists of several subtests, each measuring a somewhat different part of intelligence. The individual's performance on these various measures is summarized in five Composite scores: Verbal Comprehension, Perceptual Reasoning, Working Memory, Processing Speed and Full Scale IQ's.

For the purpose of this study, only the Perceptual Reasoning Index (PRI) subtests were used. The PRI consists of three subtests. The Block Design subtest requires the child to measure of the ability to analyze increasingly complex pictures of blocks made into geometric designs and the reproduce the designs with blocks. The Picture Concepts subtest presents several rows of pictures in which the child had to find a picture on each row that related to one another. The Matrix Reasoning subtest requires the child to choose the correct solution from several choices to increasingly complex matrices that are missing one small piece. The scores for the subtests have a mean of ten and a standard deviation of three, while the scores for the Composites have a mean of 100 and a standard deviation of 15.

WISC-IV standardization sample. The WISC-IV was normed on a sample of 2,200 children, with an equal number of males and females. March 2000, census data was matched with the standardization sample's proportion of Whites, African Americans, Hispanics, Asians and other racial groups. The sample was further stratified on age,

parent's education level, and geographic region. Also, a sample of approximately 5.7% children from special disability groups who represented the Census data of children attending school was added to the standardization sample. These groups included Mental Retardation-Mild Severity, Mental Retardation-Moderate Severity, Reading Disorder Reading and Written Expression disorders, Mathematics Disorder, Reading, Written Expression and Mathematics Disorder, Learning Disorder and Attention-Deficit/Hyperactivity Disorder, Attention-Deficit/Hyperactivity Disorder, Expressive Language Disorder, Mixed Receptive/Expressive Disorder, Open Head Injury, Closed Head Injury, Autistic Disorder, Asperger's Disorder, and Motor Impairment, and then also intellectually gifted children.

Reliability of WISC-IV. Three estimates of reliability were reported for the WISC-IV: internal consistency, test-retest stability and interscorer agreement. To measure internal consistency, the split-half method, corrected by the Spearman-Brown formula was used (except for on the Processing Speed subtests.) The overall average reliability coefficient for all age groups on the Full Scale was .97. The average coefficients for all age groups on the four Index scores were calculated as follows: Verbal Comprehension = .94; Perceptual Reasoning = .92; Working Memory = .92; and Processing Speed = .88, using the test-retest stability coefficient as a reliability estimate. Table 5 describes the reliability coefficients for the WISC-IV Indices.

Table 5

WISC-IV Indices' Reliability Coefficients

Composite/Index	Full Scale	Verbal Comprehension	Perceptual Reasoning	Working Memory	Processing Speed
Number of Subtests	n/a	3	3	2	2
WISC-IV Test-Retest Stability Coefficient	.97	.94	.92	.92	.88

Internal consistency coefficients for the individual subtests were also reported for the 16 disability groups (Wechsler, 2003, p. 34-36). Two hundred and forty-three children participated in the test-retest stability measurement of the WISC-IV. The second administrations were given 13-63 days apart, with an average interval of 32 days between administrations. Pearson's product-moment correlation was used to compute the coefficients and adequate stability over time for all age groups. The Full-Scale test-retest reliability coefficient was .93 with Index coefficients for Verbal Comprehension = .93; Perceptual Reasoning = .89; Working Memory = .89; and Processing Speed = .86. Table 6 describes the WISC-IV internal consistency coefficients.

Interscorer agreement was also measured using two independent scorers and 60 cases from the standardization sample. The coefficients ranged from .98 to .99, which the authors attributed to the basic and accurate nature of the scoring criteria.

Validity of WISC-IV. Several types of validity studies were reported for the WISC-IV, including Intercorrelational studies, Factor-Analytic studies, Cross-Validation

Table 6

WISC-IV Indices' Internal Consistency Coefficients

	Full Scale	Verbal Comprehension	Perceptual Reasoning	Working Memory	Processing Speed
WISC-IV test-retest reliability coefficient	.93	.93	.89	.89	.86

studies, and correlations with other measures (Wechsler, 2003, p. 47–74). Special group studies were also provided to add validity estimates based on test-criterion relationships (p. 75). Though no sample of D/HH children was included in the technical manual, a sample of children with Expressive Language Disorder (ELD) was included. Twenty-seven children who met the criteria for ELD, but who were still verbal, were matched with a control group were administered the WISC-IV. Individuals with ELD show lower scores on tests requiring the use of expressive language versus those requiring receptive language, and such was the case on the WISC-IV as well. Statistically significant group mean differences were found between the VCI (10.69 points) and WMI (10.58 points), while no difference was found between the ELD group and control group on the PRI and PSI. A statistically significant difference was also found on the FSIQ, 9.32 points (p. 90).

A sample of 41 students, ages six through 16 diagnosed with Mixed Receptive-Expressive Language Disorder (RELD) was also included in the technical manual (Wechsler, 2003, p. 90). Again, the RELD group was matched with a control group, and totally nonverbal children were excluded from the study. Large group differences were found for all Index scores. Children with RELD generally tended to do better on nonverbal tasks, as reflected in their WISC-IV mean score difference on the PRI of

+12.95. And, as expected the largest group differences in mean scores were found between the RELD and the control groups with differences in the FSIQ of 23.03 points, and with differences in the control group's mean VCI exceeding the RELD group's mean VCI by 21.90 points (Wechsler, p. 91).

Procedures

All participants were tested by one of the three Utah licensed school psychologists employed by the USDB. Testing was conducted for the purpose of routine eligibility and compliance psychoeducational evaluations as specified in the Individuals with Disabilities Education Act (IDEA). The participants were referred for a psychoeducational evaluation, including an assessment to measure intellectual ability. Parental consent for the evaluation was obtained, in accordance with special education laws.

At the beginning of each school year, each student's parent(s) sign a permission form stating, "On occasion, scores from student testing are gathered and used for program evaluation and research purposes, such as to investigate the use of a particular test with our populations. No identifying information, such as name, address, birth date, etc., is used so as to ensure student confidentiality." All participants selected for the study had current permission forms on file with USDB records. Refer to Appendix A for English and Spanish examples of this handout.

All three examiners received training on the administration and scoring of both instruments before the evaluations began. The three examiners have over 20 years of combined experience administering psychoeducational evaluations to D/HH students.

The UNIT and the WISC-IV were administered to 61 D/HH students across the state of Utah. In about 40% of the cases, both measures were administered during the

same testing session. However, many students' schedules required that one measure be administered on one day and the other be administered within a couple days of the first.

The UNIT was administered in accordance with standardization procedures regardless of the student's educational program (auditory/oral or sign language.) For the signing students, the WISC-IV was administered in Conceptually Accurate Signed English (CASE), which vocabulary the three examiners discussed and agreed upon before beginning to use the WISC-IV. When a sign for a word in the directions did not exist, a description of the word or word-phrase was presented in American Sign Language. For the auditory/oral students, the only modification in administration was the occasional exchange of a synonym for a word that a student obviously did not comprehend.

Because the evaluation was completed based on a referral process, all teachers and parents were informed of the results of their students through written reports, and in many cases, the results were explained during the annual Individual Education Plan (IEP) meeting.

Data Analysis

Relevant descriptive statistics were obtained for the Composite scores of the current sample of D/HH students including, means, standard deviations, ranges and skew on the WISC-IV PRI and the Standard Battery of the UNIT. A *t*-test was performed to compare the difference between the WISC-PRI scores of the current sample of D/HH students and the standardization sample of the WISC-IV PRI. Similarly, a *t*-test was performed to compare the difference between the Composite scores of UNIT Standard Battery of the current sample of D/HH students and the UNIT standardization sample. A correlation was computed to determine the comparability of scores obtained by the

current sample of D/HH students on WISC-IV PRI and UNIT. The Statistical Package for the Social Sciences (SPSS) was used to analyze the data.

Results

This study describes the cognitive scores of the D/HH students currently being served by USDB. The study provides a context for better understanding an individual's score within the continuum of the USDB D/HH WISC-IV and UNIT scores. It also addresses the difference between the WISC-IV PRI and UNIT Standard Battery Composite scores of the USDB sample and the standardization samples of each measure, and then compares the scores of USDB's sample on the two measures.

Table 7 displays the descriptive statistics for the two measures. The means, standard deviations, ranges and skew are displayed for the WISC-IV PRI scores as well as the UNIT Standard Battery Composite scores.

Table 7

Descriptive Statistics of Composite Scores for the WISC-IV PRI and UNIT Standard Battery

	<i>Mean</i>	<i>SD</i>	<i>Median</i>	<i>Range</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Skew</i>	<i>SE of Skew</i>
WISC	88.95	14.55	90.00	68	49	117	-.74	.31
UNIT	90.74	13.97	91.00	65	54	119	-.55	.31

Note. $N = 61$.

Research Question 1

What is the nature of the distribution of WISC-IV PRI scores in a robust, convenient sample of USDB D/HH students? The mean Index score of the WISC-IV PRI of the USDB D/HH sample was 88.95, the median score was 90 and the standard deviation was 14.55. The range of scores was 68, with a minimum score of 49 and a

maximum score of 117. The scores had a negative skew of $-.74$, with a standard error of $.31$. Figure 1 demonstrates the distribution.

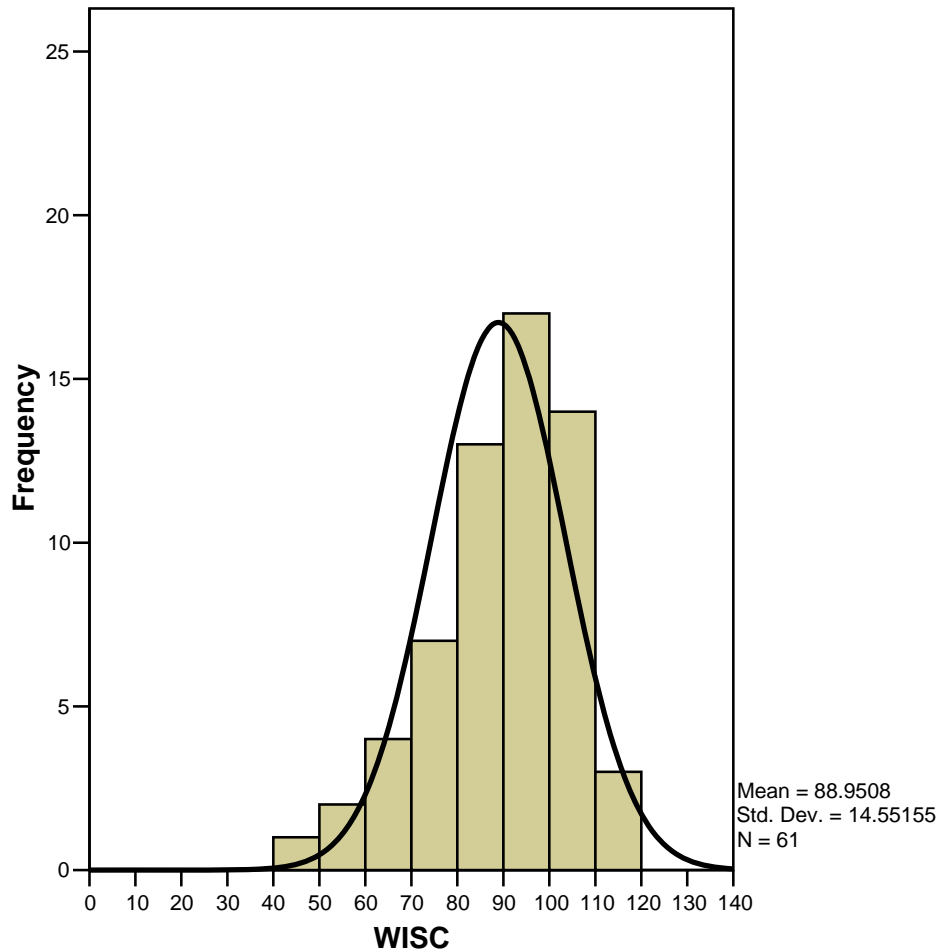


Figure 1. WISC-IV PRI Score Distribution.

Research Question 2

What is the nature of the distribution of UNIT Standard Battery Composite scores in a robust convenient sample of USDB D/HH students? The mean Composite score of the UNIT Standard Battery of the USDB D/HH sample was 90.74, the median score was

91 and the standard deviation was 13.97. The range of scores was 65 with a minimum score of 54 and a maximum score of 119. The scores had a negative skew of $-.55$, with a standard error of $.31$. Figure 2 demonstrates the distribution.

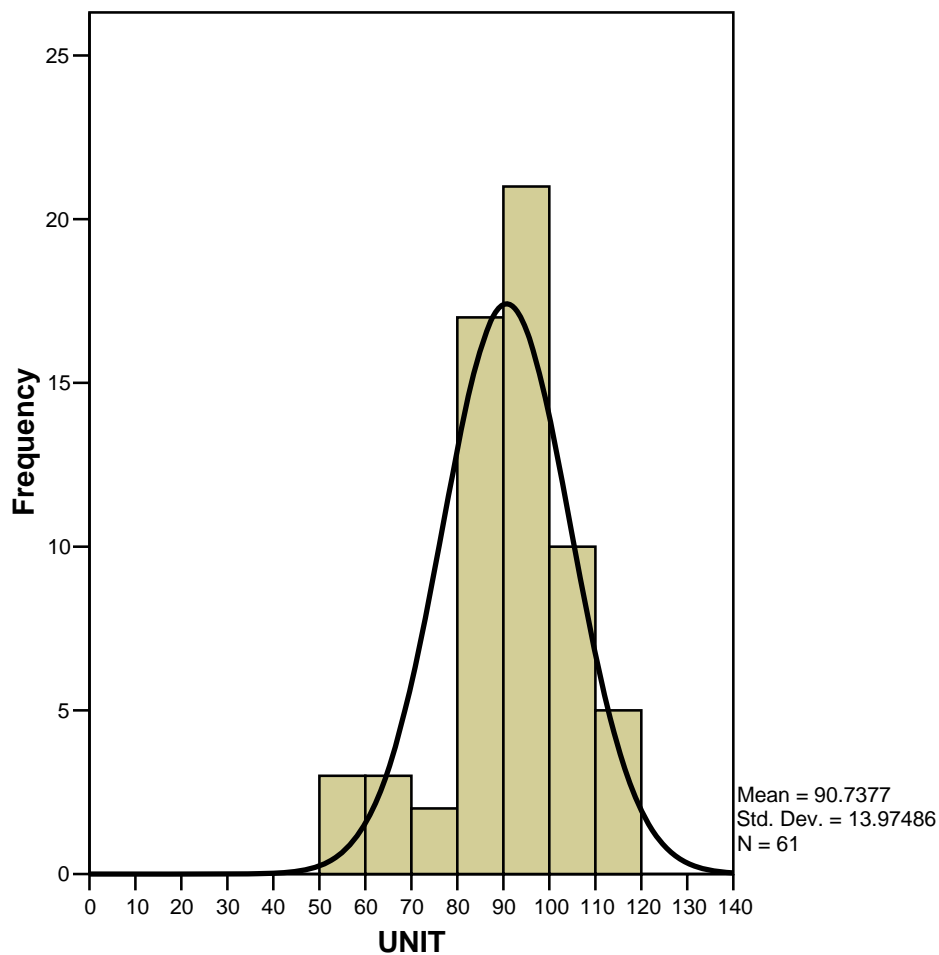


Figure 2. UNIT Standard Battery Composite Score Distribution.

Research Question 3

Is there a significant difference between a robust convenient sample of USDB D/HH students and the WISC-IV standardization sample in terms of the WISC-IV PRI scores? The mean Index score of the standardization sample on the WISC-IV PRI is 100

and the standard deviation is 15. A single sample t -test was used to determine the difference between the scores of the WISC-IV standardization sample and the USDB D/HH sample. An alpha level of .01 was identified as the cut-off for determining statistical significance in this study's data analysis. The USDB D/HH sample had a statistically significantly lower PRI score than the standardization sample of the WISC-IV, $t(60) = -5.93, p < .0001$ (two-tailed), $d = -1.05$.

Research Question 4

Is there a significant difference between a robust convenient sample of USDB D/HH students and the UNIT standardization sample in terms of the UNIT Standard Battery Composite scores? The mean Composite score of the standardization sample on the UNIT Standard Battery is 100 and the standard deviation is 15. A single sample t -test was used to determine the difference between the scores of the UNIT standardization sample and the USDB D/HH sample. An alpha level of .01 was used for significance in the statistical analysis. The USDB D/HH sample had a statistically significantly lower UNIT Standard Battery Composite score than the standardization sample of the UNIT, $t(60) = -5.18, p < .0001$ (two-tailed), $d = -9.26$.

Research Question 5

What is the correlation between the WISC-IV PRI score and the UNIT Standard Battery Composite score in a robust convenient sample of USDB D/HH? Because the distribution of USDB D/HH Composite scores on the WISC-IV PRI and UNIT Standard Battery had only slight negative skews, a Pearson correlation was performed. (However, because there was a slight negative skew, the nonparametric Spearman correlation was also performed.) An alpha level of .01 was used for significance in the statistical analysis.

The Pearson correlation coefficient was significant, $r = .75, p < .0001$ (two tailed). Thus, 56 % of the variance in one measure is explained by the other measure, indicating a high correlation between the two measures. The Spearman correlation coefficient was also significant, $r = .63, p < .0001$ (two-tailed). Figure 3 is a scatter plot of UNIT and WISC-IV PRI scores.

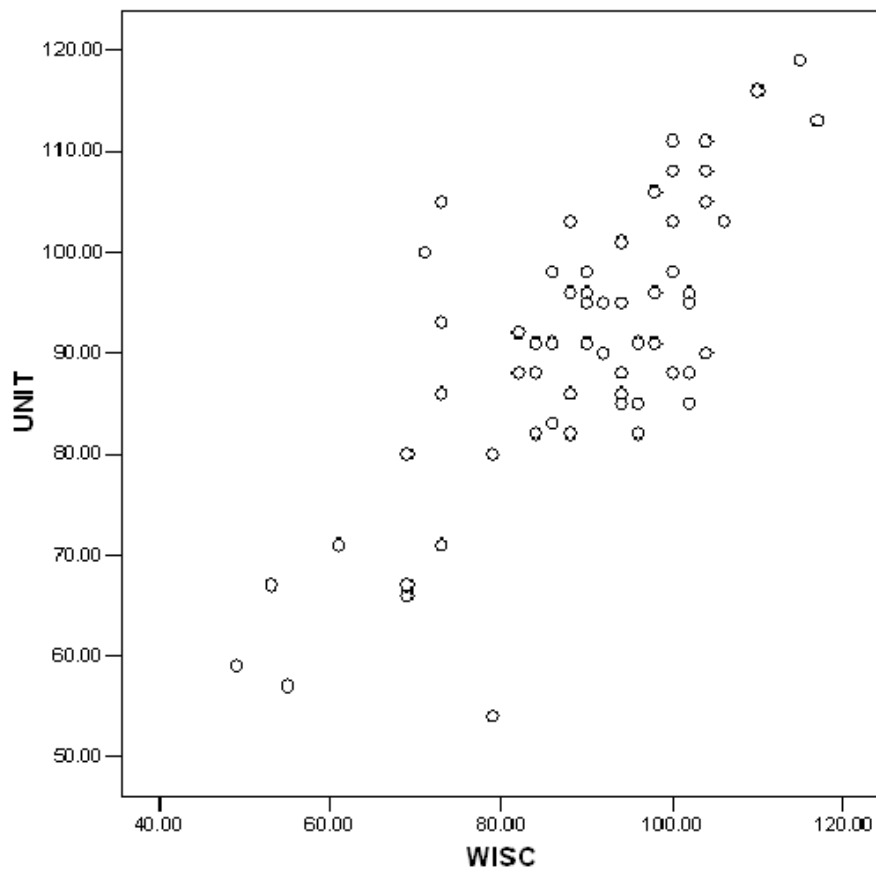


Figure 3. Scatter Plot of UNIT and WISC-IV PRI Scores.

Discussion

The major purpose of this study was to improve USDB's practice of assessing D/HH students' cognitive ability. More specifically stated, this purpose included the following goals: (a) to describe the cognitive scores of the D/HH students currently served by USDB; (b) to determine if testing time could be reduced, yet reliability of assessment results be maintained by eliminating WISC-IV assessment; and (c) to provide the Division of Services for People with Disabilities (DSPD), other state agencies, and school psychologists in Utah with the data and rationale to either support or not support the possible use of UNIT scores rather than requiring the combination of WISC-IV *and* UNIT scores.

The first purpose was accomplished with the findings that answered research questions 1, 2, 3 and 4. The result of research question one was that USDB's sample had WISC-IV PRI scores with a mean of 88.95 and standard deviation of 14.55, demonstrating a relatively normal distribution of scores. The result of research question two was that USDB's sample had UNIT Standard Battery Composite scores with a mean of 90.74 and a standard deviation of 13.97, also demonstrating a relatively normal distribution of scores. However, the means of both measures are significantly lower than the standardization sample of either measure, as both the WISC-IV PRI and the UNIT Standard Battery have means of 100. The USDB D/HH sample's mean was 11.05 points lower than the mean score of the standardization sample on the WISC-IV PRI, and the USDB D/HH sample's mean was 9.26 points lower than the mean score of the standardization sample on the UNIT Standard Battery.

Research questions 3 and 4 asked if those differences between the mean scores of the USDB D/HH sample and the standardization samples of the two measures were statistically significant. USDB's D/HH sample scores were significantly lower on both measures with a p -value of $<.0001$ in both cases.

The finding that the mean score of the USDB D/HH sample on the UNIT is approximately nine points lower than the standardization sample is very similar to the finding of the study reported in the UNIT technical manual in which the measure was administered to 106 D/HH students. The sample received a mean score that was eight points lower than the UNIT standardization sample (Bracken & McCallum, 1998, p. 192). The students in Bracken and McCallum's sample all attended a school for the D/HH and were all receiving special education, which required that the students demonstrate delayed language, lower academic performance and an inability to communicate effectively when compared with age-mate students who are not D/HH. The USDB D/HH sample seems identical in composition as the USDB D/HH students were also attending a school for the D/HH and were receiving special education. This finding seems to confirm the fact that D/HH students with educational placements similar to that of USDB do not perform as well as the standardization sample on the UNIT.

When considering the USDB D/HH UNIT scores, one of the questions posed in the statement of purpose in chapter 1 was, "What does a single USDB D/HH UNIT test score mean in the context of the total sample of USDB D/HH scores on the UNIT?" Despite the fact that the UNIT is a nonverbal measure of intelligence and was designed to provide a more fair cognitive assessment for those children disadvantaged by language loaded measures, the average USDB D/HH student has a UNIT score of 91 which actually sits within the instrument's 90% confidence interval for the *low average* range

(Bracken & McCallum, 1998, p. 262). For individuals like special educators who are used to thinking of and classifying students in terms of IQ scores, they will need to understand that the average D/HH student at USDB may look (in terms of reasoning and memory abilities,) more like a hearing student with a low average UNIT score. A D/HH student whose score is one SD below the USDB D/HH sample's mean (77) will appear to struggle compared to average D/HH students, but will appear very delayed compared to hearing peers. Similarly, a student with a score of one SD above the USDB D/HH sample's mean (105) will appear quite bright compared to average D/HH students, but will only appear average when compared with hearing peers.

Unfortunately, a comparison of the USDB D/HH sample's mean score on the WISC-IV PRI cannot be made with any other samples of D/HH children as no formal research with the WISC-IV and D/HH populations has been done. It wouldn't be particularly meaningful to compare the USDB D/HH sample's mean on the WISC-IV PRI with information from D/HH samples on the WISC-III Performance Scale because the structure of WISC-IV is significantly different (Braden 2005) than the WISC-III and the two composites have only one subtest in common.

WISC-IV PRI scores do not add any information to UNIT scores, and the question posed above can again be answered, namely, "What does a single USDB D/HH WISC-IV PRI test score mean in the context of the total sample of USDB D/HH scores on the WISC-IV PRI?" The answer to this version of the question is not notably different from the answer presented for UNIT. The average D/HH student at USDB may look (in terms of reasoning and memory abilities,) more like a hearing student with a low average UNIT score. A D/HH student whose score is one SD below the USDB D/HH sample's mean (74) will appear to struggle compared to average D/HH students, but will appear

very delayed compared to hearing peers. Similarly, a student with a score of one SD above the USDB D/HH sample's mean (104) will appear quite bright compared to average D/HH students, but will only appear average when compared with hearing peers.

The 5th and final research question was answered by showing that there is a strong correlation between the WISC-IV PRI scores and the UNIT Standard Battery Composite scores among the USDB sample. Since the scores on the two measures are so highly correlated, this study effectively provides data and rationale to support the use of the only one of the two instruments when a measure of the cognitive abilities of D/HH children is required.

One outcome of using only one measure, instead of both measures, is to reduce by half the amount of academic time students miss because of intelligence testing. Further, the school psychologists who administer the tests save even more time because not only do they save the time it takes to administer the second test, but they are also saved the time it takes to score it and to write the report. If the school psychologists at USDB saved only two hours of time for each of the 61 participants in this study, more than three full weeks of work would be saved. Since there is a statewide shortage of school psychologists, three weeks of school psychology time devoted to other activities could really make a difference to a school and/or district.

If only one measure is going to be used, the logical measure to choose is the UNIT for several reasons. The first reason is that the authors considered D/HH children when creating the norms for the instrument. They included D/HH children in the standardization sample. And, the instructions can be administered to D/HH children whether they use Auditory/Oral, Total Communication or American Sign Language to communicate as the instructions are presented completely nonverbally in gestures.

Therefore, after considering all the reasons above, the study indicates that there is not any significant information to be gained by administering the WISC-IV in addition to the UNIT.

Another reason to choose the UNIT is that in addition to the Standard Battery Composite score that it provides, it has four other Composite scores including, Memory, Reasoning, Symbolic and Nonsymbolic, as opposed to the one composite score provided by the WISC-IV PRI. While the Composite scores for the UNIT are each based on only two subtests, they do provide a basis of comparison in different areas of cognitive processing thus allowing practitioners the ability to obtain a better overall picture of the D/HH student's nonverbal functioning.

Limitations and Future Research

The sample of USDB D/HH students used in this study was convenient. Only about a third of the D/HH students attending USDB were included. This means that the age of the students, the preferred communication method, the educational program, and the level of hearing loss were not considered as separate variables, even though any or all of the variables could contribute to differences in cognitive abilities.

Another limitation of this study is that as 75% of the D/HH students in Utah do not attend self-contained classes through USDB (though almost all receive Outreach language enrichment services,) it is difficult to generalize this finding to all D/HH students. Because the students at USDB tend to require more intensive service than those who receive services in their neighborhood schools, it is possible that the sample of D/HH students included in this study does not represent those D/HH students in Utah who do not require intensive, self-contained special education services. Further research that included a sample of D/HH students receiving services in their neighborhood schools

would have to be conducted in order to adequately describe the IQ distribution of all D/HH students in the state of Utah.

Braden's (2005) recommendation that the WISC-IV be "systematically studied to evaluate its performance with deaf and hard-of-hearing" (p. 372) was accomplished in part with the current study, as the less verbal WISC-IV PRI was found to correlate with the nonverbal UNIT. But, further research on the use of the WISC-IV with D/HH children could be undertaken. For instance, the Processing Speed subtests and Index scores and the Working Memory subtest and Index scores could provide useful information about the functioning of D/HH children if their performance on the WISC-IV was evaluated and published. However, in both cases it is nearly impossible to administer the subtests in the standardized format to a D/HH child who uses sign language, which negates the validity of their scores.

Alternatively, for D/HH children using the Auditory/Oral method of communication who are mainstreamed or considering a mainstream placement, verbal intelligence scores, such as the WISC-IV Verbal Comprehension subtests and Index scores, could provide a comparison between the verbal abilities of the D/HH student and his/her hearing peers. However, as this study's findings show that the scores on the UNIT are comparable to WISC-IV PRI scores, but the WISC-IV can not be administered to D/HH children in a standardized manner, it seems when a nonverbal intelligence score is needed, the more appropriate recommendation would be to leave the WISC-IV out of the assessment process for D/HH children.

Conclusions and Recommendations

Historically, measures of intelligence with evidence of validity and reliable scores have not been created and published specifically for D/HH individuals. As a result, it is

current practice in Utah school districts and social service agencies to request Wechsler scores whenever an intelligence score is needed. This practice does not take into account the fact that D/HH individuals are disadvantaged by a verbally administered measure such as the Wechsler. However, a completely nonverbal cognitive measure, the UNIT is available for use and its authors took D/HH children into account when creating it.

The results of this study are tied to one and only one major conclusion: School districts and social service agencies should discontinue the use of the WISC-IV with D/HH children. The test of preference when assessing cognitive ability must be the UNIT. The UNIT is an instrument already familiar to most school psychologists and can be administered without accommodations. On the other hand, the WISC-IV subtests require that the examiner have knowledge about D/HH children in order to make the necessary accommodations to be able administer the assessment. This requires additional skill and training on the part of the examiner.

This study indicated that the composite scores USDB D/HH children received on the WISC-IV PRI and the UNIT Standard Battery were highly correlated. And as such, only the UNIT need be administered. The UNIT is a more appropriate measure than the WISC-IV PRI of the cognitive functioning of D/HH children because of the inclusion of D/HH children in the standardization sample, the use of completely nonverbal and gestural administration, and comparative breadth of information the UNIT provides. A simple handout could be given to school district and social service personnel to explain the benefits of UNIT scores over WISC-IV PRI scores for the cognitive assessment of D/HH children. Refer to Appendix B for an example of this handout.

References

- Allen, T. (1986). *Understanding the scores: Hearing-impaired students and the Stanford Achievement Test* (7th ed.). Washington, DC: Gallaudet University Press.
- Blennerhassett, L. (1990). Intellectual assessment. In D. F. Moores & K. P. Meadow-Orlans (Eds.), *Educational and developmental aspects of deafness* (pp. 255-280). Washington, DC: Gallaudet University Press.
- Blennerhassett, L., & Traxler, C. B. (1999). WISC-III utilization with deaf and hard of hearing students. *Gallaudet Research Institute Technical Report 99-1*. Washington, DC: Gallaudet University Press.
- Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test* (UNIT). Itasca, IL: Riverside Publishing.
- Braden, J. P. (1992). Intellectual assessment of deaf and hard-of-hearing people: A quantitative and qualitative research synthesis. *School Psychology Review, 21*(1), 82-94.
- Braden, J. P. (1994). *Deafness, deprivation, and IQ*. New York: Plenum 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. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation scientist-practitioner perspectives* (pp. 351-376). Boston, MA: Elsevier Academic Press.
- Braden, J. P., & Athanasiou, M. S. (2005). A comparative review of nonverbal measures of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 557-577). New York: Guilford Press.

- Braden, J. P., Maller, S. J., & Paquin, M. M. (1993). The effects of residential versus day placement on the performance IQs of children with hearing impairment. *The Journal of Special Education, 26*(4), 423-443.
- Brown, L., Sherbenou, R. J., & Johnson, S. K. (1997). *Test of Nonverbal Intelligence—Third Edition (TONI-3)*. Austin, TX: Pro-Ed.
- Brown, S.C. (1986). Eitological trends, characteristics, and distributions. In A. N. Schildroth & M.A. Karchmer (Eds.), *Deaf children in America* (pp. 33-54). San Diego, CA: College Hill.
- Cash, R. M. (1994). *An analysis of the performance of students who are hearing impaired on selected traditional measures of psychoeducational assessment*. Unpublished doctoral dissertation, Brigham Young University, Provo, UT. Retrieved October 6, 2007, from *ProQuest Digital Dissertations* database. (Publication No. AAT 9507258)
- Chovan, W., & Benfield, J. R. (1994). Varied sign language systems and their mediating effects on WISC-R verbal subtests of profoundly deaf students: A replication. *Perceptual and Motor Skills, 78*, 61-62.
- Cone-Wesson, B. (2003). Screening and assessment of hearing loss in infants. In M. Marschark & P. E. Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (pp. 21-37). New York: Oxford University Press.
- England, C. T. (1997). Predictive and concurrent validity of the *Universal Nonverbal Intelligence Test* and the *Test of Nonverbal Intelligence-II* with achievement. *Dissertation Abstracts International* (UMI No. 9823082).
- Fagan, T. K. (2002). School psychology: Recent descriptions, continued expansion, and an ongoing paradox. *School Psychology Review, 31*, 5-10.

- Flanagan, D. P., & Harrison, P. L. (2005). *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.). New York: Guilford Press.
- Gallaudet Research Institute. (2003). *2001-2002 regional and national survey*. Washington, DC: Author. Retrieved October 28, 2007, from <http://gri.gallaudet.edu/demographics/>
- Hammill, D. D., Pearson, N. A., & Wiederholt, J. L. (1996). *Comprehensive Test of Nonverbal Intelligence*. Austin, TX: Pro-Ed.
- Hooper, V. S. (2002). *Concurrent and predictive validity of the Universal Nonverbal Intelligence Test and the Leiter International Performance Scale - Revised*. Doctoral dissertation, The University of Tennessee, Knoxville. Retrieved October 6, 2007, from *ProQuest Digital Dissertations* database. (Publication No. AAT 3075546)
- Hughes, D., Sapp, G. L., & Kohler, M. P. (2006). *Issues in the intellectual assessment of hearing impaired children*. University of Alabama, Birmingham. (ERIC document reproduction service No. ED493819). Retrieved June 17, 2007, from http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/f1/a4.pdf
- Hunt, H. L. (1997). *Assessing intelligence in deaf children: Is the verbal scale of the WISC-III valid?* Doctoral dissertation, Gallaudet University, District of Columbia. Retrieved October 6, 2007, from *ProQuest Digital Dissertations* database. (Publication No. AAT 9823067).
- Karchmer, M. A., & Mitchell, R. E. (2003). Demographic and achievement characteristics of deaf and hard-of-hearing students. In M. Marschark & P. E.

- Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (pp. 21-37). New York: Oxford University Press.
- Klein, D. H., & Parker, E. W. (2002). *Spoken communication for students who are deaf or hard of hearing; a multidisciplinary approach*. Portland, OR: Butte Publications.
- Knooks, H., & Vervloed, M. P. J. (2003). Educational programming for deaf children with multiple disabilities: Accommodating special needs. In M. Marschark & P. E. Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (pp. 451-463). New York: Oxford University Press.
- Krivitski, E. C. (2000). *Profile analysis of deaf children using the Universal Nonverbal Intelligence Test*. Doctoral dissertation, State University of New York at Albany. Retrieved October 6, 2007, from *ProQuest Digital Dissertations* database. (Publication No. AAT 9979509).
- Mackinson, J. A. (1996). *The validity of the Test of Nonverbal Intelligence-2 with deaf children*. Doctoral dissertation, Gallaudet University, District of Columbia. Retrieved October 6, 2007, from *ProQuest Digital Dissertations* database. (Publication No. AAT 9712692).
- Maller, S. J. (1997). Deafness and WISC-III item difficulty: Invariance and fit. *Journal of School Psychology, 35*(3), 299-314.
- 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. (2000). Item invariance in four subtests of the *Universal Nonverbal Intelligence Test* (UNIT) across groups of deaf and hearing children. *Journal of Psychoeducational Assessment, 18*, 240-254.
- Maller, S. J. (2003). Intellectual assessment of deaf people: A critical review of core concepts and issues. In M. Marschark & P. E. Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (pp. 451-463). New York: Oxford University Press.
- Maller, S. J. (2005). Review of the *Wechsler Intelligence Scale for Children-4th ed.* In R. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* (pp. 1093-1096). Lincoln, NE: University of Nebraska 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: WISC-III Monograph, 105-113*.
- Maller, S. J., & Ferron, J. (1997). WISC-III factor invariance across deaf and standardization samples. *Educational and Psychological Measurement, 57*(6), 987-994.
- Maller, S. J., & French, B. F. (2004). Universal Nonverbal Intelligence Test factor invariance across deaf and standardization samples. *Educational and Psychological Measurement, 64*, 647-660.
- 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: Oxford University Press.
- Marschark, M., Lang, H. G., & Albertini, J. A. (2002). *Educating deaf students: From research to practice*. New York: Oxford University Press.

- Mather, N., & Wendling, B. J. (2005). Linking cognitive assessment results to academic interventions for students with learning disabilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 269-294). New York: Guilford Press.
- McCallum, R. S., & Bracken, B. A. (1997). The *Universal Nonverbal Intelligence Test*. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 268-280). New York: Guilford Press.
- McCallum, R. S., Bracken, R., & Wasserman, J. (2001). *Essentials of nonverbal assessment*. New York: John Wiley and Sons.
- Mitchell, R. (in press). How many deaf people are there in the United States? Estimates from the survey of income and program participation. *Journal of Deaf Studies & Deaf Education*.
- Moore, D.F. (1996). *Educating the deaf*. Boston, MA: Houghton Mifflin Co.
- Mykelbust, H. R. (1953). Towards a new understanding of the deaf child. *American Annals of the Deaf*, 98, 345-257.
- Mykelbust, H. R. (1960). *The psychology of deafness*. New York: Grune & Stratton.
- Pinter, R., & Patterson, D. (1917). A comparison of deaf and hearing children in visual memory for digits. *Journal of Experimental Psychology*, 2, 76-88.
- Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition*. Itasca, IL: Riverside.
- Roid, G. H., & Miller, L. J. (1997). *Leiter International Intelligence Scale—Revised*. Wood Dale, IL: Stoelting.

- Slate, J. R., & Fawcett, J. (1996). Gender differences in Wechsler performance scores of school-age children who are deaf or hard of hearing. *American Annals of the Deaf, 141*(1), 19-23.
- Sullivan, P. M., & Montoya, L. A. (1997). Factor analysis of the WISC-III with deaf and hard-of-hearing children. *Psychological Assessment, 9*(3), 317-321.
- Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review, 38*, 406-427.
- United States Department of Education. (2002). *Outcome data for students with low-incidence disabilities*. (20 U.S.C. 1462 § 662 c 3).
- 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. 3-38). New York: Guilford Press.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—3rd edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—4th edition*. San Antonio, TX: The Psychological Corporation.

APPENDIX A

Parent/Guardian Permission Forms



Utah Schools for the Deaf and the Blind

Annual Student Permission Form

Student Name _____

INSTRUCTIONS: Please read each of the following paragraphs. Indicate your desires and instructions by checking the appropriate spaces. Finally, sign and date on lines provided on the reverse side of this form.

1. All students are encouraged to participate in all school activities, including lessons/tutorials with Utah Schools for the Deaf and the Blind (USDB) outreach teachers and Related Services (RS) staff, field trips and excursions as a part of their education. *My child has my permission to participate in all school-sponsored lessons, fieldtrips, excursions and extracurricular activities, whether at a school facility or away, provided that supervision by the USDB faculty and/or staff is provided.*

_____ YES _____ NO _____ Does Not Apply

2. On occasion, children may be photographed/video-recorded to show student progress and/or while participating in school activities by staff, university training programs, news media, or for public relations information use by USDB. *USDB has my permission to photograph, video record and/or audio record my child which may be shown to individuals or groups for training, information, displays, and public relations purposes.*

_____ YES _____ NO _____ Does Not Apply

3. USDB receives requests for address lists of students and parents from parent teacher organizations, non-profit service organizations, and other public agencies. *USDB has my permission to release my name and address to non-profit organizations and agencies which in the administration's view will benefit me or my child.*

_____ YES _____ NO _____ Does Not Apply

4. On occasion, USDB RS providers must transport students from their schools to the administrative offices for testing purposes. *USDB RS providers have my permission to transport my child to and from school in an authorized state vehicle for testing purposes.*

_____ YES _____ NO _____ Does Not Apply

5. USDB participates in Medicaid School-Based Development services in accordance with IDEA, Part B. This participation is to defray the costs of special education and related services provided to children enrolled at USDB and enrolled in Medicaid. *USDB has my permission to claim Medicaid reimbursement on behalf of my child for school-based skills development services.*

_____ YES _____ NO _____ Does Not Apply

(See Reverse Side)

6. On occasion, scores from student testing are gathered and used for program evaluation and research purposes, such as to investigate the use of a particular test with our populations. No identifying information, such as name, address, birth date, etc., is used so as to ensure student confidentiality. *I give permission for my student's test scores to be used for program evaluation/research purposes, and I understand that no identifying information will be attached to any of the test scores.*

_____ YES _____ NO _____ Does Not Apply

7. On occasion, students need to or wish to leave school property during school hours, or to use transportation other than that provided by USDB. This may be beneficial as students become older and/or improve their abilities to become independent. Such activities may also be a part of visually impaired students training in orientation and mobility. *My child has my permission to (check the appropriate statements):*

Leave the school grounds without USDB adult supervision.

_____ YES _____ NO _____ Does Not Apply

Leave the school grounds if he/she has been designated as an independent traveler by an orientation and mobility specialist.

_____ YES _____ NO _____ Does Not Apply

Leave the school grounds with another student who has been approved by an Orientation and Mobility Specialist as a sighted guide.

_____ YES _____ NO _____ Does Not Apply

Leave the school with individuals other than USDB staff and in vehicles not provided by USDB.

_____ YES _____ NO _____ Does Not Apply

If you answered yes to any of the above, *please list names and/or titles of individuals with whom your child may leave school grounds (such as Orientation and Mobility instructors).*

I have read the above information and indicated my desires for each item. I understand that I may withdraw my permission at any time by notifying the Utah Schools for the Deaf and the Blind in writing.

Parent (Guardian) Signature

Date



Utah Schools for the Deaf and the Blind

Planilla anual de permiso del alumno

Nombre del alumno

INSTRUCCIONES: Por favor lea cada uno de los siguientes párrafos. Indique lo que desea y las instrucciones chequeando los espacios apropiados. Finalmente, firme y ponga la fecha en las líneas en la parte de atrás de la planilla.

8. Todos los alumnos estas animados a participaren todas las actividades de la escuela, incluyendo lecciones/tutorías con el personal de Utah Schools for the Deaf and the Blind (USDB) maestros de afuera y de servicios relacionados, paseos y excursiones como parte de su educación. *Mi niño/a tiene mi permiso a participar en todas las lecciones patrocinadas por la escuela, paseos, excursiones y actividades extracurriculares, sean en las facilidades de la escuela o afuera. Previendo que supervisión por personal de la facultad de USDB sea proveída.*

SI NO No se aplica

9. En ocasiones los alumnos pueden ser fotografiados/ grabados por el personal, entrenadores de programa de la universidad, noticias o por información de relaciones publicas usada s pro USDB para ver el progreso del alumno y/o mientras participan en actividades de la escuela .*USDB tiene mi permiso para fotografiar y grabar mi niño/a lo cual pueden mostrar a individuos o grupos por información de entrenamiento y propósito de relaciones publicas.*

SI NO No se aplica

10. A USDB organizaciones de padres y maestros, organizaciones de servicio sin ganancias y otras agencias públicas le piden listas de direcciones de alumnos y padres. *USDB tiene mi permiso de dar mi nombre y dirección a organizaciones sin ganancias y agencias en las cuales la administración vea beneficioso para mi niño/a.*

SI NO No se aplica

11. En ocasiones, proveedores RS de deben transportar alumnos de la escuela a las oficinas administrativas para hacerles exámenes. Los proveedores RS de USDB tienen mi permiso para transportar a mi hijo/a con propósito de exámenes de y para la escuela en un vehiculo autorizado por el estado..

SI NO No se aplica

12. USDB participa en servicios basados en desarrollo de escuela de Medicaid de acuerdo a IDEA, parte B. Esta participación es para descarrilar el costo de educación especial y servicios relacionados proveídos a alumnos de USDB que son parte de Medicaid. *USDB tiene mi permiso de pedir reembolso a Medicaid en parte de mi hijo/a para servicios de desarrollo de habilidades basadas en la escuela.*

SI NO No se aplica

(Vea el lado de atrás)

13. En ocasiones, notas de los exámenes de los alumnos se toman y se usan para evaluación del programa y propósitos de investigación, como el uso de un examen particular en nuestra población. No se usa información de identificación como nombre, dirección, fecha de nacimiento etc., para mantener la confidencialidad del alumno. Doy permiso para que los resultados de exámenes de mi niño/a se usen para evaluación del programa /propósitos de evaluación/propósitos de investigación y entiendo que información de identificación no va a estar atada con los resultados de exámenes..

SI NO No se aplica

14. En ocasiones los alumnos tienen que salir de la escuela durante horas escolares o usar transportación que no es proveída por USDB. Esto puede traer beneficios para los alumnos cuando son mayores y/o mejoren sus habilidades de volverse independiente. Actividades que pueden ser parte de niños con impedimento visual entrenando en orientación y movilidad. Mi niño/a tiene mi permiso de (marque lo apropiado):

Salir de la escuela sin supervisión de adulto de USDB.

SI NO No se aplica

Salir de la escuela si el especialista de orientación y movilidad le ha asignado un compañero independiente.

SI NO No se aplica

Salir de la escuela con otro alumno que sea aprobado como guía de vista por un especialista de movilidad y orientación.

SI NO No se aplica

Salir de la escuela con personas que no sean parte del personal de USDB y en vehículos que no sean de USDB.

SI NO No se aplica

Si contesto sí a cualquiera de las preguntas de arriba, por favor escriba los nombres de las personas con quien su niño/a pueda salir de la escuela. (Como instructores de orientación y movilidad).

Leí y entendí la información de arriba e indique mis deseos en cada punto. Entiendo que puedo quitar mi permiso en cualquier momento notificando por escrito Utah Schools for the Deaf and the Blind.

Firma del Padre (Guardián)

Fecha

APPENDIX B

Handout for Utah Schools Serving USDB Students



Utah Schools for the Deaf and the Blind

Challenges and Recommendations for Intelligence Testing of Deaf and Hard of Hearing (D/HH) Students

Challenges in administering intelligence tests to D/HH students:

- Wechsler scales such as the WISC-IV are the most commonly administered intelligence tests to students in the U.S., including D/HH students.
- Traditional intelligence tests are administered verbally, posing an inherent disadvantage to the D/HH, because hearing loss typically results in delayed language development.
- D/HH students have a difficult time understanding overall directions and information related to specific test items.
- D/HH students struggle to respond in an age-appropriate manner.
- The instructions of verbal tests do not easily translate into sign language so individuals who communicate with sign language cannot adequately access the information they need in order to respond appropriately.
- Verbal subtests are not valid measures of intelligence in D/HH students, whether the students are in Auditory/Oral, Total Communication or American Sign Language programs.
- Although traditional intelligence tests have subtests that appear to be nonverbal, the subtests are not truly non-verbal, but are instead “language reduced.”
- Traditional intelligence tests’ standardized administration must be modified for D/HH students, which largely invalidates test results.
- Traditional intelligence tests do not include D/HH students in their standardization samples, which largely invalidates test results.

Recommendations for administering intelligence tests to D/HH students:

- The UNIT has a completely nonverbal administration method.
- The UNIT has six subtests that provide reasoning, memory, symbolic and nonsymbolic quotients, as well as a full scale score.
- The UNIT included D/HH students in the standardization sample.
- The UNIT included a sample of 106 D/HH students in a post-standardization validity study: D/HH students were found to have a mean Full Scale score of 92, compared to the mean Full Scale score of 100 for the standardization sample.
- A sample of 61 D/HH students at USDB had a mean UNIT Full Scale score of 91, which was highly correlated with its mean Perceptual Reasoning Index score on the WISC-IV of 89.
 - The *average* D/HH student may look (in terms of reasoning and memory abilities) more like a hearing student with a *low average* UNIT score.
 - A D/HH student whose score is one SD below the USDB D/HH sample’s mean (74) will appear to struggle academically compared to average D/HH students, but will appear very delayed compared to hearing peers.
 - A student with a score of one SD above the USDB D/HH sample’s mean (104) will appear quite bright compared to average D/HH students, but will only appear average when compared to hearing peers.
- **The UNIT appears to be a valid, reliable measure of intelligence for D/HH students.**

References available on back

References

- Bracken, B. A., & McCallum, R.S. (1998). *Universal Nonverbal Intelligence Test (UNIT)*. Itasca, IL: Riverside Publishing.
- Braden, J. P. (1992). Intellectual assessment of Deaf and Hard-of-Hearing people: A quantitative and qualitative research synthesis. *School Psychology Review*, 21(1), 82-94.
- Braden, J. P. (1994). *Deafness, deprivation, and IQ*. New York: Plenum 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. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation scientist-practitioner perspectives* (pp. 351-376). Boston, MA: Elsevier Academic Press.
- Braden, J. P., & Athanasiou, M. S. (2005). A comparative review of nonverbal measures of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 557-577). New York: Guilford Press.
- Krivitski, E. C. (2000). *Profile analysis of deaf children using the Universal Nonverbal Intelligence Test*. Unpublished doctoral dissertation, State University of New York, Albany. Retrieved October 6, 2007, from ProQuest Digital Dissertations database. (Publication No. AAT 9979509)
- Maller, S. J. (2000). Item invariance in four subtests of the *Universal Nonverbal Intelligence Test (UNIT)* across groups of deaf and hearing children. *Journal of Psychoeducational Assessment*, 18, 240-254.
- Maller, S. J. (2003). Intellectual assessment of deaf people: A critical review of core concepts and issues. In M. Marschark & P. E. Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (pp. 451-463). New York: Oxford University Press.
- Maller, S. J. (2005). Review of the *Wechsler Intelligence Scale for Children- 4th ed.* In R. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* (pp. 1093-1096). Lincoln, NE: University of Nebraska Press.
- 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: Oxford University Press.
- Marschark, M. (2006). Intellectual functioning of deaf adults and children: Answers and questions. *European Journal of Cognitive Psychology*, 18(1), 70-89.
- Voorhies, L. (2007). *Existing practice and proposed changes in cognitive assessment of Utah students identified as deaf and hard of hearing*. Unpublished doctoral dissertation, Brigham Young University, Provo, UT.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—4th edition*. San Antonio, TX: Psychological Corporation.