View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Philadelphia College of Osteopathic Medicine: DigitalCommons@PCOM

Philadelphia College of Osteopathic Medicine

DigitalCommons@PCOM

PCOM Psychology Dissertations

Student Dissertations, Theses and Papers

2020

The Equivalency of Digital and Paper-based Administration of the Wechsler Intellegence Scale for Children-Fifth Edition with Deaf and Hard-of-hearing Students

Kenneth Reimer

Philadelphia College of Osteopathic Medicine

Follow this and additional works at: https://digitalcommons.pcom.edu/psychology_dissertations

Part of the School Psychology Commons

Recommended Citation

Reimer, Kenneth, "The Equivalency of Digital and Paper-based Administration of the Wechsler Intellegence Scale for Children-Fifth Edition with Deaf and Hard-of-hearing Students" (2020). *PCOM Psychology Dissertations*. 537.

https://digitalcommons.pcom.edu/psychology_dissertations/537

This Dissertation is brought to you for free and open access by the Student Dissertations, Theses and Papers at DigitalCommons@PCOM. It has been accepted for inclusion in PCOM Psychology Dissertations by an authorized administrator of DigitalCommons@PCOM. For more information, please contact library@pcom.edu.

Philadelphia College of Osteopathic Medicine

School of Professional and Applied Psychology

Department of School Psychology

THE EQUIVALENCY OF DIGITAL AND PAPER-BASED ADMINISTRATION OF THE WECHSLER INTELLEGENCE SCALE FOR CHILDREN-FIFTH EDITION WITH DEAF AND HARD-OF-HEARING STUDENTS

By Kenneth Reimer

Submitted in Partial Fulfillment of the Requirements for the

Degree of Doctor of Psychology

March 2020

PCOM SCHOOL OF PROFESSIONAL AND APPLIED PSYCHOLOGY™

DISSERTATION APPROVAL

This is to certify that the thesis presented to us by Kenneth Reimer on the 2nd day of March, 2020, in partial fulfillment of the requirements for the degree of Doctor of Psychology, has been examined and is acceptable in both scholarship and literary quality.

COMMITTEE MEMBERS' SIGNATURES

Chairperson

Chair, Department of School Psychology

Dean, School of Professional & Applied Psychology

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ABSTRACT

This study explored the equivalence of the digital and paper-based version of the Wechsler Intelligence Scale for Children, 5th Edition (WISC-V) with deaf and hard-ofhearing children. There are a number of psychological tests being administered using both paper and digital formats. There is currently no literature on the validity of using these new digital tests with deaf and hard-of-hearing children. This study used a repeated measures design in which deaf and hard-of-hearing participants took the Figure Weights and Matrix Reasoning subtests of the WISC-V in counterbalanced order. The study found no format effect for the Figure Weights subtest with this population. However, a small format effect was found with this population on the Matrix Reasoning subtest. This study found that there are some differences of performance within this population when administering the WISC-V using different formats. Additional research is needed to confirm the findings of this study and expand to include additional subtests as well as additional IQ tests.

CHAPTER 1: INTRODUCTION

According to the most recent data released by the federal government, over 6.5 million students in the United States currently are labeled as having specific disabilities (Cortiella & Horowitz, 2014). This means that roughly 13% of the students in a typical school have disabilities. First enacted in 1975, the Education for All Handicapped Children Act, now known as the Individuals with Disabilities Education Act, established the mandate for public schools to provide services for students with disabilities

(Education for All Handicapped Children Act, 1975). This legislation placed a burden on public schools to both identify and provide services to students with disabilities. The field of school psychology flourished following the new mandate to identify students. Psychological testing became one of the most common methods for identifying students in need of special education services, and it can help families and school professionals better understand students' strengths and weaknesses. The results of testing are used to make many important decisions, including those regarding access to special-education services, determination of disabilities, and qualifying for gifted programs.

The Wechsler Intelligence Scales for Children-Fifth Edition is one of the many tests of intelligence available (WISC-V) and used with a wide variety of populations (Wechsler, 2014). This test has traditionally been administered using an easel with test items printed on paper. A student indicates his or her response to the stimulus printed on a page, and the examiner records the response on a paper protocol. With the advancement of technology, this process is now also available in an electronic medium in which an iPad screen takes the place of the traditional easel. The evolution to an electronic medium has streamlined administration, automating scoring and accelerating the pace of administration. With the transition from the WISC-IV to the WISC-V, some school systems are choosing to purchase the digital version of the WISC-V in lieu of the print version.

The Gallaudet Research Institute (2013) estimated that there are over 45,000 deaf and hard-of-hearing school-aged children in the United States. The deaf and hard-ofhearing population requires additional considerations with the use of the WISC-V. There are no specialized intelligence tests designed solely for use with deaf and hard-of-hearing students, but many studies have provided support for the use of a variety of intelligence tests with this population (e.g., J. P. Braden, 1992; Pick, 2013; Sullivan & M. Vernon, 1979).

Statement of the Problem

The new digital format has streamlined the administration of the WISC-V as compared to the paper-based version in several different ways. The time saved and reduction of scoring errors make using the digital format more attractive to practitioners. Given that many students depend on an accurate administration and interpretation of intelligence tests such as the WISC-V to qualify for special education services, it is essential that the digital version work as reliably as the paper version. Regrettably, deaf and hard-of-hearing students were not included in the equivalency study between the paper and digital formats of the WISC-V (Daniel, Wahlstrom, & Zhang, 2014). Before the digital format can be used confidently with deaf and hard-of-hearing students, further studies are needed to ensure equivalency between the traditional and digital versions of the WISC-V.

On the digital format of the WISC-V, a student is given an iPad on which he or she is able to view the stimulus as well as touch his or her responses. Student engagement, familiarity of technology, and motivation may be influenced differently when given the paper or digital format of the test. For instance, a previous study found that scores on the Matrix Reasoning and Picture Concepts subtests were higher on the digital administration (Daniel, 2012). The study did not explain this performance difference between the paper and digital formats of the test. This is especially alarming given that fluid reasoning subtests are typically the least culturally loaded subtests on the WISC-V and, thus, the best predictors of intelligence in individuals who are deaf or hardof-hearing (J. P. Braden, 1992). Due to the varied results across equivalency studies, there is reason to believe that differences in performance may exist on some composites of the WISC-V (Daniel et al., 2014). Recently, several studies have demonstrated that the digital version of the WISC-V is equivalent in individuals with autism spectrum disorder and accompanying language impairment, attention-deficit/hyperactivity disorder (ADHD), intellectual giftedness, and intellectual disabilities (Raiford, Drozdick, & Zhang, 2015; Raiford, Holdnack, Drozdick, & Zhang, 2014). A study was needed to test whether the digital system would provide equivalent scores to the paper-based test for individuals who are deaf or hard-of-hearing. Because the fluid reasoning subtests are the most appropriate to measure intelligence in individuals who are deaf or hard-of-hearing due to their low language demands, the specific subtests that were investigated included Matrix Reasoning and Figure Weights.

Purpose of the Study

Given that there is documented equivalency of the paper-based and digital administrations of the WISC-V on hearing students, the purpose of this study was to determine whether raw scores are equivalent between paper-based and digital administrations of the WISC-V for students who are identified as either deaf or hard-ofhearing. Despite studies supporting the use of digital administration with special populations, no studies on the use of this format currently exist examining the deaf or hard-of-hearing population. Since practitioners are currently using both the paper and digital formats of the WISC-V, it is essential to know whether students would obtain similar scores regardless of format used. If equivalence is demonstrated, the norms, reliability, and validity information gathered for the paper format can be applied to the digital results for this population, while keeping in mind that deaf and hard-of-hearing students were not part of the normative sample.

Definitions of Terms

The *Individuals with Disabilities Education Act* (IDEA; 2014) is a law that makes a free and appropriate public education available to all children, including those with disabilities. *Special education* is individualized instruction provided for a student identified with a disability (Education for All Handicapped Children Act, 1975). *Psychological Testing* is detailed evaluation of a child's strengths and weaknesses in several areas, such as cognitive, academic, language, behavioral, emotional, and social functioning (Sattler, 2008). The *Wechsler Intelligence Scales for Children-Fifth Edition (WISC-V)* is an individually administered, comprehensive clinical instrument for assessing the intelligence of children (Wechsler, 2014). The *Matrix Reasoning Subtest* is a part of the WISC-V in which the child views an incomplete matrix or series and selects the response option that completes the matrix or series (Wechsler, 2014). The *Picture Concepts Subtest* is a part of the WISC-V in which a child views two or three rows of pictures and selects one picture from each row to form a group with a common characteristic (Wechsler, 2014). *Digital administration* refers to the process of administering a psychological assessment with the use of an iPad. *Fluid reasoning is* a part of cognitive functioning involving a broad pattern of reasoning including seriation, sorting, and classifying (Horn & Blankson, 2005). A *normative sample* is a group of children included in the development of an assessment tool that reflect the performance of the population as a whole (Sattler, 2008). The term *deaf and hard-of-hearing* refers to a label which can be applied to individuals with a hearing loss of more than 20 decibels (National Association of the Deaf, 2016).

CHAPTER 2: LITERATURE REVIEW

The History of Psychological Testing

Psychological testing has played an important role in assessing mental abilities throughout history. Psychological assessments are tests of maximal performance, which ask individuals questions or have them perform tasks to their best abilities (Sattler, 2008). These tests are designed to categorize and compare performance along a host of dimensions, such as memory, attention, executive functioning, visual processing, verbal reasoning, and processing speed. These tests generally fall into two groups: cognitive tests that measure the potential of one's processing and achievement assessments that measure the amount of knowledge one has obtained.

Psychological testing is an always evolving science that has made advancements in understanding human development thanks to many contributors from around the world. The first recorded use of tests to sort individuals for employment and other classifications can be traced back to ancient China in 2200 B.C. (Wainer,1988). These tests determined who was eligible to obtain employment with the government and other coveted positions of power. These tests were objective in nature and great lengths were undertaken to remove examiner bias from the testing process. The next major advancement in standardized testing came in 1599, when standardized rules were established for exams, many of which are still used in modern times (McGucken, 2008). In the past two centuries, France adopted similar testing practices in the 1790s, Britain in the 1830s, and Massachusetts in the 1860s (Wainer, 1988). The United States Congress endorsed similar tests by implementing the Civil Service Act in 1883 (Theriault, 2003).

The goal of this legislation was to ensure government employees were hired on the basis of their skills and knowledge and to avoid nepotism and corruption. The next major development in psychological assessment was advanced by Francis Galton. Galton's test of sensory and motor skills was developed in the late 1800s, but his largest contribution to modern psychological testing was his statistical work demonstrating that a normal distribution could be applied to any human attribute, including those measured by his sensory and motor test. Galton opened a center to the public where individuals could undergo a series of assessments and receive the written results, which is the foundation of today's modern psychological report. One of the greatest criticisms of Galton's work was his stance that intelligence varied between populations, suggesting one's intelligence was in part determined by nationality. His conclusions resulted from his work with hereditary genes and eugenics, which has been widely debunked (Redvaldsen, 2017).

The next major evolution of psychologist testing was thanks in large part to the work of Wilhelm Wundt. Wundt was a German psychologist whose work focused primarily on attention span, perception, and reaction time (Wundt, Creighton, & Titchener, 1894). His work laid the foundation for the Stanford-Binet Scales, which were the first psychological tests to correlate performance on a task with a mental age for that skill (U.S. Army, 1918). Around the same time, James Cattell, an American psychologist and professor, was conducting research using a series of tests on college students. This research consisted of 10 tests, including Dynamometer Pressure, Rate of Movement, Sensation-Areas, Pressure Causing Pain, Least Noticeable Difference in Weight, Reaction-Time for Sound, Time for Naming Colors, Bi-Section of a 50-Centemeter Line, Judgement of 10 Seconds Time, and Number of Letters Remembered on Once Hearing (J. Cattell, 1890). This set of tests has gone through multiple iterations and became known as the Wechsler Scales, which have become the most widely used intelligence tests in the United States in modern times (Sattler, 2008).

Although there have been many contributors to modern psychological testing, Spearman's role in psychometrics research is arguably one of the most critical to the proliferation of psychological testing. His major contribution of the use of reliability coefficients allowed for the results of psychological tests to be used for estimation and predictions. Additionally, Spearman (1927) put forth the current theory that intelligence has a shared variance across sets of cognitive tasks. In other words, he proposed that there is a global intelligence underlying all cognitive tasks (Dearborn, 1927). As many researchers were developing tests that could be administered individually, Thorndike, Thurstone, and Otis created tests that could be scored with a key, allowing for completely objective measures along with group administrations (Morgan & Steinman, 1943).

As the United States entered the First World War, Otis's work became the foundation for the Army Alpha and Beta tests thanks in large part to the advocacy of the American Psychological Association. These tests were originally designed for use with adults to improve selection and occupational placement during the First World War

(Terman, 1918). These tests could be adapted for group use, correlated with measures of intelligence, used to measure a wide range of abilities, and had quick and objective scoring, included alternative forms, discouraged malingering and cheating, were not reliant on school training, minimized written responses, and were quick to administer (DuBois, 1970). During this time, the Army Alpha test included Oral Direction, Arithmetical Reasoning, Practical Judgement, Synonym-Antonym, Disarranged Sentences, Number Series Completion, Analogies, and Information. The Beta test included Incomplete-Picture and Coding tasks that were timed (Yerkes, 1921). The Beta test was designed to be used with illiterate and non-English speaking men, allowing for a wider group of individuals to be screened by the United States Military. Combined, the Alpha and Beta tests were given to an estimated two million men. During the Second World War, the United States furthered its testing capabilities through the development of the Army General Classification Test. This test consisted of four subtests, Reading and Vocabulary, Arithmetic Computation, Arithmetic Reasoning, and Spatial Relations (U.S. Army, War Department, 1941). This test allowed for the prediction of a telegrapher's speed as well as success in a wide range of military tasks (Wainer, 1988).

Building on the Army Alpha and Beta, Wechsler and Bellevue designed a general test of cognitive abilities called the Wechsler-Bellevue Intelligence Scale (Wechsler, 1939a). The Wechsler-Bellevue Intelligence Scale went on to lay the groundwork for the Wechsler Intelligence Scale for Children (Wechsler, 1949). This test was the first mainstream psychological test designed specifically for children and featured both a verbal and performance scale. Since then, four major revisions of this psychological test have been developed. As psychological testing has become more common since the 1970s, it has also become increasingly used with children and adolescences for a range of purposes. Psychological testing is used for screening of disabilities, giftedness, and neuropsychological conditions such as ADHD (Sattler 2008).

Today, researchers are continuing the long tradition by developing new tests and extending the role of psychologist testing. Tests have expanded to include focuses on development, learning and memory, attention, and achievement. Modern tests have evolved to reflect current theories of intelligence and started to include normative samples with diverse populations. Some popular tests have even been translated into languages other than English. Although the speed of testing has exploded in the past few decades, the contributions of early researchers laid a solid foundation for contemporary psychological tests.

Theories of Intelligence

Varied theories of intelligence make up the theoretical foundations underlying the development of psychological tests. Because intelligence is not a single construct that is agreed upon in the literature, it is important to understand the history and differences between the major theories of intelligence that were used in the development of modern psychological instruments.

Jean Esquirol was first cited to make a distinction between individuals who had mental illness and those who never developed their intellectual capacities (Huertas, 2008). As far back as 1890, Boas and Gilber used sensorimotor tests to categorize students as either “bright” or “dull.” (J. S. Braden, DiMarino-Linnen, & Good, 2001). These early researchers were laying the foundation for classifying individuals based on their mental abilities. In 1905, the Binet-Simon Scale became the first psychological test designed to be used with children to diagnose mental retardation and became the model for future tests (Sattler, 2008). Since its inception, the Stanford Binet scale has become arguably the most important tool in helping to identify students who have learning difficulties or need special education services (Sattler, 2008).

During the early 20th century, there were two opposing theories of intelligence held by Binet and Goddard. Goddard's belief that there was a single underlying function of intelligence determined by heredity was the leading theory at the time, whereas Binet viewed intelligence as more malleable due to environmental factors, although still related to genetics (Terman, 1919). Stern (1914) defined the mental quotient as a mental age divided by chronological age, which was then multiplied by 100. The 1916 revision of the Stanford-Binet was updated by Terman and changed Stern's mental quotient to the term intelligence quotient. Terman's classifications were based on the percentage of children who passed at each age level and the items resulted in a median intelligence quotient of 100 (Stern, 1914). Robert Yerkes advocated strongly against the age-scale format. He believed that test items should measure the same construct throughout development, which was referred to as the point-scale format. One of the major criticisms of this method was that partial credit was given for partial answers. Additionally, his method did not produce the same degree of brightness, so results could not be compared across age ranges (Otis, 1917). The third revision of the Stanford-Binet Scale included updated norms and the use of standard scores in place of the previous ratio intelligence quotient (Terman & Merrill, 1973). In 1939, David Wechsler also adopted the point-scale format of intelligence testing when he adapted existing tests into the Wechsler-Bellevue Intelligence Scale (Wechsler, 1939b). Wechsler considered intelligence to be global in nature and part of an individual's personality. His work attempted to measure effective intelligence in contrast to Thurston's work that attempted to measure primary abilities.

There are two major milestones in defining intelligence in the field of psychology: the 1921 and 1986 symposiums. Both symposiums included adaptations to the environment, basic mental processes, and higher-order thinking; however, the 1986 symposium broadened the definition to include metacognition and executive processes (Sternberg & Detterman, 1988). As the field of psychology evolves, so does the definition of intelligence. In addition, culture plays a role in the definition of intelligence. For example, it is important to note that it is a Western cultural tradition to celebrate problem-solving and logic. In contrast, in Eastern cultures, it is more common to prioritize social intelligence and identifying contradictions (Sattler 2008). Many of the contemporary definitions emphasize the ability to adjust to the environment, ability to learn, and to perform abstract thinking (Sattler, 2008 & Wechsler, 1958). Modern researchers fell into two camps in regard to their views of intelligence. Spearman, Vernon, and Carroll viewed intelligence as a general and specific factor (g, s), whereas Thorndike, Thurstone, Guilford (1967), Cattell, and Horn subscribed to a multifactor theory of intelligence (Sattler, 2008).

Edward Thorndike put forth the multifactor theory of intelligence, which stated that intelligence is made up of interconnected but distinct intellectual abilities. Specifically, his definition of mental abilities fell into three clusters: social intelligence, concrete intelligence, and abstract intelligence (Thorndike & Columbia University, 1927). Thurstone used centroid factor analysis, which led to seven primary ability factors, all with equal weight: verbal comprehension, word fluency, number skills, memory, perceptual speed, inductive reasoning, and spatial visualization (Thurstone, 1938).

Guildford developed a three-dimensional structure of intelligence that included operations, content, and product (Guildford, 1967). His model expanded previous work to include 125 possible factors of intelligence. The next major evolution in the theory of intelligence came from the work of Raymond Cattell and John Horn, which put forth the notion that intelligence fell into fluid and crystallized intelligence (Horn & R. Cattell, 1966). Fluid intelligence was defined as nonverbal, culture-free mental efficiency, whereas crystallized intelligence was defined as acquired skills and knowledge that depended on exposure to culture. Horn's theory of intelligence has evolved over time to now include 87 primary mental abilities and 8 second-order abilities, including acculturation knowledge, fluid reasoning, short-term memory, long-term memory, processing speed, visual processing, auditory processing, and quantitative knowledge (Horn & Blankson, 2005).

In contrast to the multifactor theories of intelligence, the general and specific factor theorists viewed intelligence as a two-factor theory. Charles Spearman's theory put forth the idea that a general factor, (g), was the general mental energy that was required by a task and more difficult tasks had a high (g) loading (Spearman, 1927). Similarly, Philip Veron's theory of hierarchical intelligence included a (g) factor, in addition to two major group factors below, verbal-educational and spatial-mechanical (P. E. Vernon, 1950). John Carroll proposed a three-stratum factor analytic theory of cognitive abilities, including major group factors, minor group factors, and specific factors (Carroll, 1993). Carroll's eight broad factors consisted of fluid intelligence, crystallized intelligence, general memory and learning, broad visual perception, broad auditory perception, broad retrieval ability, broad cognitive speediness, and processing speed.

Currently, both Spearman's and Thurstone's views on intelligence are widely accepted, with many practitioners falling along a continuum between their two theories to define intelligence. Given that intelligence is thought to consist of individual subskills in addition to a global (g) intelligence, modern researchers have started to measure intelligence in new ways. These tests of intelligence use a variety of tasks that measure cognitive reasoning skills.

Although giving a single comprehensive battery to students is commonplace, Flanagan, Ortiz, and Alfonso (2013) developed a cross-battery assessment style. Psychologists use this method, which entails selecting subtests from different intelligence and neuropsychological assessments to examine a child's cognitive abilities, rather than administering a complete single intelligence test. This has permitted examiners to take advantage of the best parts of different assessments to collect information about an individual's cognitive abilities. The development of the crossbattery method has impacted the way in which psychologists administer tests, which involves using parts of different tests to assess the abilities of an individual. The crossbattery method permitted examiners to use selected subtests to measure specific constructs, such as verbal or fluid reasoning. The cross-battery approach is a time efficient method to measure cognitive abilities in a more flexible way than giving one intelligence test. Additionally, the cross-battery approach allows for the assessment of cognitive strengths and weaknesses in individuals from culturally and linguistically diverse backgrounds, such as students who are deaf and hard-of-hearing (Flanagan, Ortiz, & Alfonso, 2013).

The Foundation of Testing in Children

The Wechsler Intelligence Scale for Children (WISC) was first published in 1949 and contained 11 of the subtests from the Wechsler-Bellevue Intelligence Scale adapted for use with children ages 6 years, 0 months to 16 years, 11 months old (Wechsler, 1939a, 1949). The 11 subtests resulted in a Full Scale IQ (FSIQ), a Verbal IQ and a Performance IQ. The Performance IQ on this test forms the foundation of modern fluid reasoning and nonverbal measures. The Wechsler Intelligence Scale for ChildrenRevised (WISC-R) expanded the age range and maintained the same subtests and indexes as the original test (Wechsler, 1974). The next revision of the test, the Wechsler Intelligence Scale for Children-Third Edition (WISC-3) added one new subtest and reorganized the indexes into the Verbal Comprehension Index, Perceptual Organization Index, the Freedom from Distractibility Index, and the Processing Speed Index (Wechsler, 1991). The most recent predecessor to the current edition, the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) was published in 2003 and contained several revisions (Wechsler, 2003). The WISC-IV eliminated the Verbal IQ and Performance IQ scores and retained the remaining 10 subtests from the WISC-III. The WISC-IV organized the 10 subtests into the Verbal Comprehension Composite, the Perceptual Reasoning Composite, the Working Memory Composite, and the Processing Speed Composite (Wechsler, 2004). Additionally, the Perceptual Reasoning Index and Working Memory Index were used for the first time in the WISC-IV.

The Wechsler Intelligence Scale for Children, Fifth Edition

Wechsler defined intelligence as the capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment (Wechsler 1944). He believed that his tests measured several of the key parts of intelligence while knowing that one test could not measure all aspects of intelligence. In fact, Wechsler believed that intelligence tests actually measure an individual's resourcefulness to cope with challenges.

The most recent edition, the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V), was published in 2014 and contains a total of 21 subtests and 13 index scores. The newest edition contains five primary indexes: Verbal Comprehension Index, Visual Spatial Index, Fluid Reasoning Index, Working Memory Index, and Processing Speed Index. Several ancillary indices are also included, such as the Quantitative Reasoning Index, Auditory Working Memory Index, Nonverbal Index, General Ability Index, and Cognitive Proficiency Index, which can provide additional information regarding a child's cognitive abilities (Wechsler, 2014). Lastly, the WISC-V also includes the Naming Speed Index (NSI), Symbol Translation Index (STI), and Storage and Retrieval Index (SRI), which are considered complementary index scores and are designed to provide information based on clinical need (Wechsler, 2014). Thirteen subtests were retained from the WISC-IV: Block Design, Similarities, Matrix Reasoning, Digit Span, Coding, Vocabulary, Symbol Search, Information, Picture Concepts, Letter-Number Sequencing, Cancellation, Comprehension, and Arithmetic. The WISC-V added the Figure Weights, Visual Puzzles, Picture Span, Naming Speed Literacy, Naming Speed Quantity, Immediate Symbol Translation, Delayed Symbol Translation, and Recognition Symbol Translations subtests. These subtests may be administered in isolation or as part of a complete battery. Visual Puzzles and Figure Weights were adapted from the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV; Wechsler, 2008) and Picture Span was adapted from the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition (WPPSI-IV; Wechsler, 2012). Word

Reasoning and Picture Completion subtests were dropped from the most recent revision.

The WISC-V is an individually-administered, comprehensive clinical instrument for assessing the intelligence of children ages 6 years, 0 months through 16 years, 11 months (Wechsler, 2014). The WISC-V was developed for and normed on children that were raised in the United States. Students who were born outside the United States or who live in homes where English is not the primary language spoken may face a disadvantage on some of the verbal subtests (Sattler, 2008). The WISC-V measures both broad intellectual functioning and discrete cognitive domains. The WISC-V is highly correlated with intellectual functioning as measured on the Differential Ability ScalesSecond Edition (Elliott, 2007thehe Kaufman Assessment Battery for Children, Second

Edition (Kaufman & Kaufman, 2004), the Stanford-Binet Intelligence Scales, Fifth Edition (Roid, 2003), and the Woodcock-Johnson IV (Mather & Jaffe, 2016).

The WISC-V test kit contains all of the needed material to administer the test battery. The WISC-V Administration and Scoring Manual contains all instructions required to administer all subtests and complete the Summary and Primary Analysis pages of the Record Form. There are three Stimulus Books, which contain the subtests and are bound for easy use. Two of the Stimulus Books contain the needed stimuli for the primary and secondary subtests used to derive the Primary Index scores, the FSIQ, and all Ancillary Index scores. The third Stimulus Book contains the stimuli for the complementary subtests that derive the NSI, STI, and SRI. There is also an Administration Supplement that contains information to complete the Ancillary Index, Complementary Analysis, and Processing Analysis page of the Record Form.

Changes were made from the WISC-IV to substantially reduce the number of items needed to meet the discontinue rule. For example, Similarities was reduced from 5 consecutive scores of 0 on the WISC-IV to 3 consecutive scores of 0 on the WISCV. Also, the discontinue rule was standardized to 3 for all primary subtests for consistency.

There are 11 types of recordable errors and 6 types of process observations available while scoring the WISC-V. In regard to errors, the examiner can document Block Design Dimension Error, Block Design Rotation Error, Coding Rotation Error, Symbol Search Set Error, Symbol Search Rotation Error, Naming Speed Literacy Error, Naming Speed Color-Object Error, Naming Speed Size-Color-Object Error, Naming Speed Letter-Number Error, and Naming Speed Quantity Error. In regard to process observations, the examiner can document Don't Know, No Response, Item Repetition, Requested Repetition, Subvocalization, and Self-Corrections.

The normative sample for the WISC-V was obtained from April of 2013 to March of 2014 and chosen based on several demographic variables: age, sex, race/ethnicity, parent education level, and geographic region. This sample was matched to the October 2012 U.S. census data (Wechsler, 2014).

Interpretation of the WISC-V

The WISC-V has four levels of interpretation, including Full Scale, Primary Index, Ancillary Index, and Complementary Index (Wechsler, 2014). The traditional interpretation of a the WISC-V includes administration of seven subtests that make up the FSIQ. The FSIQ is the most reliable and psychometrically sound score when measuring a typical child's intellectual functioning with the WISC-V. Often, a child's intellectual functioning will be summarized using this global score. A clinician can interpret beyond the FSIQ by looking at the child's performance on the primary index scale level. At this level, the seven subtests that comprise the FSIQ are classified in five domains: Verbal Comprehension, Visual Spatial, Fluid Reasoning, Working Memory, and Processing Speed. The Primary Index Scales are represented by standard scores, similar to the FSIQ. The third level of interpretation is the use of the five ancillary index scale levels. A child's performance can be broken down into the Quantitative Reasoning Index, Auditory Working Memory Index, Nonverbal Index, General Ability Index, and Cognitive Proficiency Index. The ancillary index scores are derived from both primary subtests that make up the FSIQ and primary index scales, and secondary subtests to provide more information about a child's functioning in these areas. The fourth way that a clinician can interpret a child's performance on the WISC-V is at the complementary index scale level. This includes the NSI, STI, and SRI.

A child's performance on the subtest level is measured by totaling the raw scores, or the total points a child earns on a task. Notably, raw scores are not age-corrected, meaning they do not provide enough information and need to be converted to a scale that allows for performance to be compared to same aged peers. On the WISC-V subtests, a child's raw scores is converted to scaled scores, which have a mean of 10 and a standard deviation of 3. These subtests' scaled scores are combined into composite scores, which have a mean of 100 and a standard deviation of 15.

Although the FSIQ is often the most reliable and valid measure of a child's cognitive functioning on the WISC-V, there are times when the use of this score is inappropriate. Deaf and hard-of-hearing students are one subgroup in which the traditional use of the FSIQ is often not a valid measure of underlying cognitive functioning (National Association of School Psychologists [NASP], 2012). The FSIQ includes the use of the Verbal Comprehension Index, which is one of the most culturally loaded indexes on the WISC-V. On the WISC-V, interpretation of performance on the Similarities, Vocabulary, Information, and Comprehension subtests are problematic for deaf and hard-of-hearing children (Day, Adams Costa, & Raiford, 2015). For these subtests, there may be an underlying assumption that the deaf or hard-of-hearing child has had similar access to the content of these subtests. Additionally, by translating verbal items into equivalent signs or fingerspelling, the words may significantly modify the subtest items. Lastly, there are no scoring guidelines for signed responses, which negatively impacts scoring reliability (Day et al., 2015).

Although the FSIQ is often not a valid method of interpreting a deaf or hard-ofhearing child's performance on the WISC-V, a child's performance can be interpreted using the second most psychometrically sound method of using the primary index scales. One of the most useful primary index scales for use with deaf and hard-of-hearing children is the Fluid Reasoning Index because it is the least culturally loaded index on the WISC-V (Wechsler, 2014). The WISC-V eliminated the Perceptual Reasoning Index and replaced it with the Visual Spatial Index and Fluid Reasoning Index. The Fluid Reasoning Index is designed to measure reasoning while minimizing the impact of language, making it ideal for assessing deaf and hard-of-hearing students' cognitive abilities. In fact, deaf and hard-of-hearing individuals have been found to have similar performance to their hearing peers on performance measures (McCallum, 2017). The Fluid Reasoning subtests of the WISC-V have been shown to have a strong g-loading for overall intelligence (Brue & Wilmshurst, 2016). Given that Fluid Reasoning subtests have been shown to be a strong indicator of overall intelligence, this provides the clinician the ability to make reliable interpretations of cognition in deaf and hard-ofhearing students using the Fluid Reasoning Index.

There are five subtests on the WISC-V that makes up the Nonverbal Index (NVI), which includes subtests that do not require expressive responses. “The NVI offers a more appropriate estimate of overall ability for children with substantial expressive language delays or other clinical conditions with expressive verbal difficulties” (Wechsler, 2014, 34). The NVI is also used to estimate overall ability for children who are deaf or hard-of-hearing, as well as those who are English language learners. It is critical to discuss the exclusionary criteria for the children included in the normative sample of the WISC-V. The lack of deaf and hard-of-hearing students in the normative sample impacts the test’s validity when used with this population. In the WISC-V norming, students were excluded if their primary languages were not English, they were primarily nonverbal or uncommunicative, they had disruptive behaviors or insufficient compliance with testing, they were tested with any intelligence measures in the previous 6 months, they had uncorrected visual impairments, they had an uncorrected hearing loss, they had upper extremity disabilities that would affect motor performance, or they were previously or currently diagnosed with any physical conditions, neurological conditions, psychological conditions, or illnesses that might depress test performance (Wechsler, 2014). The normative sample did, however, include students with specific educational classifications: developmental delays, intellectual disabilities, specific learning disabilities, speech/language impairments, ADHD, and gifted and talented. One major criticism of this effort is the failure to match the percentage of students in the normative sample with the percentage of individuals with these classifications seen in the U.S. population.

Lastly, the WISC-V demonstrated test-retest reliability for subtests, processes, composites, and complementary scores. The range of time for the test-retest intervals was 9 to 82 days with a mean of 26 days. Overall, the WISC-V demonstrated adequate stability across time for all age ranges of the test. Vocabulary was excellent (.90), Similarities, Information, Comprehension, Block Design, Visual Puzzles, Figure Weights, Arithmetic, Digit Span, Picture Span, Letter-Number Sequencing, Coding, Symbol Search, and Cancellation were good (all approximately .80), and Matrix Reasoning and Picture Concepts were acceptable (both approximately .70). It should be noted that the stability coefficient for the Fluid Reasoning Index is the lowest off all of the indexes of the WISC-V. The large amount of guessing on this index may be contributing to this lower reliability.

Fluid Reasoning Subtests

The updated version of the WISC-V has separated Visual Spatial subtests from Fluid Reasoning subtests based on factor analysis results. The new Fluid Reasoning Index includes Matrix Reasoning and Figure Weights, with Picture Concepts and Arithmetic as substitutes. There are 32 items on the Matrix Reasoning subtest and 34 items on the Figure Weights subtest. Each response on an item in both subtests results in a raw score of 1 for a correct response and a 0 for an incorrect response. According to the WISC-V manual, on the Matrix Reasoning subtest “the child views an incomplete matrix, or series, and selects the response option that completes the matrix or series. The task requires the child to use visual-spatial information to identify the underlying conceptual rules that link all the stimuli and then apply the underlying concepts to select the correct response” (Wechsler, 2014, 9). The Manual goes on to describe the Figure Weights subtest in detail:

> Within a specified time-limit, the child views a scale with missing weight(s) and selects the response option that keeps the scale balanced. This task requires the child to apply the quantitative concept of equality to understand the relationship among objects and apply the concepts of matching, addition, and/or multiplication to identify the correct response. (Wechsler, 2014, 9).

Due to the structure and task demands of Matrix Reasoning and Figures Weights, both subtests can be completed with minimal language except for when explaining the instructions (Day et al., 2015). On the Picture Concepts subtest, “a child views two or three rows of pictures and selects one picture from each row to form a group with a common characteristic” (Wechsler, 2014, 9. This subtest is designed to measure fluid and inductive reasoning, visual-perceptual recognition, and conceptual thinking. There are 27 items on the subtest, 7 of which are new for the WISC-V. On the Arithmetic subtest, a child “mentally solves arithmetic problems within a specific time limit” (Wechsler, 2014, 9). This subtest involves mental manipulation, concentration, attention, working memory, and numerical reasoning. There are 34 items on this subtest, both visual and verbal, 18 of which were substantially modified from the WISC-IV. According to the WISC-V Technical and Interpretive Manual, the Fluid Reasoning subtests have moderate correlation with one another and to the Verbal Comprehension, Visual Spatial, and Working Memory subtests (Wechsler, 2014). This is likely due to the use of multiple components of executive functioning as well as the high *g*-loading on these tasks. These subtests provide a multiple-choice response format.

Digital Administration of the WISC-V

Administering cognitive assessments is typically one of the most time-consuming responsibilities that school psychologists have throughout the school year. School psychologists have strict timelines to which they must adhere when children are referred for an evaluation (IDEA, 2004). Although each state is able to set its own timeline for testing a student after he or she has been referred for special education, the number of students referred for testing can create a backlog. School psychologists are trained to provide a wide range of valuable psychological services, including counseling, consultation, and interventions. Often, testing will take up time that could be spent providing these other needed services. One study found that the average school psychologist will spend 50% of his or her work hours engaged in testing activities (Stinnett, Havey, & Oehler-Stinnett, 1994). In addition to the amount of time a school psychologist spends on testing, time is also devoted to scoring, report writing, and presenting results to teachers and parents. The use of the digital administration provides timesaving over the paper administration, due to automatic scoring, freeing up valuable time for practitioners. The consequences of this work are not trivial; the results of testing can have profound ramifications for a child's education. Eligibility for special education is often influenced by the results of psychological testing. Obtaining meaningfully low or high scores are sometimes required to qualify for specific services under categories such as intellectual disability or gifted and talented (U.S. Department of Education, 2001). Therefore, student testing is a critical role that school psychologists must execute as efficiently as possible. Additionally, given the profound impact testing results have on a child, the results of the assessment must be trustworthy.

Q-Interactive

According to Pearson, the company that publishes the WISC-V, the Q-interactive system was designed to make assessment more convenient and accurate. The Qinteractive system is a comprehensive digital system that is used to administer and score tests that are traditionally given using paper-based tests. Testing takes place on two iPads in an app called Assess. The test administrator uses an iPad to access the test administration instructions, score and record responses, and control visual stimuli while the examinee uses a second iPad to view and respond to stimuli. The tests are loaded onto the iPads from a website called Central, where clients are set up and reports are generated. The tests can be selected on either the Assess application or Central website and then sent to the iPad. The Q-interactive system uses two iPads that are synced via Bluetooth and allows an examinee to select his or her response by touching the selection on a screen. The examinee's response is automatically recorded and then double-checked by the examiner. This removes the element of having to hand score each item, which reduces scoring mistakes. In addition to reducing the opportunity for mistakes, the digital version's physical format streamlines the administration process. For example, the stimulus of the digital test is presented on the screen of an iPad, which reduces the need to flip a page for each question item. And the use of the iPad itself removes the need of having to carry the stimulus books to the testing location. These changes in the digital format provide a much needed convenience for time constrained school psychologists. With the digital version of the test, the school psychologist does not need to carry cumbersome paper protocols, a stopwatch, or pencil, as these functions are built into the digital version of the test. All of these features make the digital version on the Qinteractive system more convenient than the traditional paper testing kits.

Nevertheless, even with these improvements in the efficiency of administering the test, the examiner must ensure that the test is valid. Ensuring that a test has good psychometric properties is essential to obtaining valid and reliable results (Sattler, 2008). Raw score equivalency between the digital and paper version of the WISC-V has been achieved across several studies (Daniel et al., 2014; Raiford et al., 2015; Raidford et al., 2014). These studies found that the number of items answered correctly were equivalent between the two formats of the WISC-V. Since testing can have a profound impact on a student's education, the paper-based and digital versions of the test must produce similar results when given to an individual. A student's performance on the test cannot be influenced by the format of the test, otherwise a student may not obtain a valid score.

One interesting finding from equivalency studies is that on the WISC-IV Matrix Reasoning and Picture Concepts subtests, children performed better when completing tasks that involved conceptual reasoning with detailed visual stimuli on the tablet when compared to the paper tests (Daniel et al., 2014). This study revealed that the Matrix Reasoning and Picture Concept subtests from the WISC-IV had effect sizes above the .20 cutoff score established for the study. This means that the students performed better on the digital administration of the WISC-V on these two subtests compared to the paperbased administration. These were the only two subtests that were significantly different. These findings are troubling given that the two subtests from the WISC-V that measure fluid reasoning might be influenced by the format of the test. These differences were not investigated further, leaving no explanation as to why students would perform better on the digital administration than the paper-based administration on these subtests.

Although these studies had many good qualities such as equivalent-groups reliability, as well as test-retest reliability, the selection of participants is a second design flaw of the Daniel et al.'s 2014 study. In the study, the researchers screened out children with perceptual disabilities, motor disabilities, and other clinical populations. Psychologists often use these tools with clinical populations, who have characteristics excluded from this study (Fiorello, 2007). In other words, students who are given these tests in an educational setting are suspected of having disabilities, and such students were not included in the equivalency study. The current research does not provide a rationale for the differences in performance between the two formats. Lastly, no children who were deaf or hard-of-hearing were included in this study. Given the importance of these subtests in measuring the cognitive abilities in this clinical population, it is critical that the score be reliable between the paper and digital versions of this test.

The Use of Q-Interactive for Psychological Assessments

The use of the Q-interactive platform has expanded to include the WISC-V, the WISC-V Spanish Edition, the WAIS-IV, and the WPPSI-IV. In addition to cognitive assessments, Q-interactive also now includes many achievement, executive function, speech and language, memory, and neuropsychology assessments.

One of the most compelling reasons to use the Q-interactive platform over the paper format of a test is the increased engagement of the individuals being assessed. In one study, a vast majority of Q-interactive practitioners reported observing an effect of the Q-interactive on the children’s level of engagement (Daniel, 2013). The use of Qinteractive with clinical groups such as students with autism, ADHD, intellectual disabilities, learning disabilities, or developmental delays were also supported. The findings suggested that Q-interactive increased examinees’ engagement and attention, which was most observable in younger children, ages 5 through 9, when compared to older children, ages 10 through 18 (Daniel, 2013).

The use of the Q-interactive platform has several additional advantages over the paper-based kit including accuracy, portability, efficiency, flexibility, and focus on the examinee (Weiss, Saklofske, Holdnack, & Prifitera, 2015). The Q-interactive platform provides the advantage of automating the subtest rules such as start points, stop points, and discontinuations. The iPads also allow for greater portability, taking the place of the physical test kits. The software scores examinees' responses, making the overall administration and scoring time more efficient. The digital test provides greater flexibility by allowing the addition of subtests from various batteries based on performance during the assessment.

Although there are many advantages to using the digital administration of tests using the Q-interactive system, there are some drawbacks as well. First, the test has to be transmitted to the iPads using a Wi-Fi connection or the iPads must be connected to WiFi in order to load assessments into the Assess application. Given that a psychologist gives tests in multiple settings, it requires advanced preparation and planning to ensure the appropriate tests are loaded onto the device. If the examiner decides to change or add subtests while testing a student, he or she must ensure a Wi-Fi connection is available to purchase or load additional subtests. Second, the order in which a psychologist administers the subtests must be established before the testing session begins. The examiner is unable to skip a subtest after the administration has started. Third, the digital administration limits the capability of an examiner to test the limits, which is helpful in gathering qualitative information on a student's performance (Sattler, 2008). Once an examiner has given a subtest to a client, he or she is unable to go back into the subtest to re-administer any items for additional information.

Q-interactive has its own unique workflow in order to administer a psychological assessment battery. The Central is a web-based portal where a practitioner can create client folders, select assessments, rearrange the administration order, and set the time and date of the assessment. The assessment is then sent to the Assess application located on the practitioner's iPad. The test is then accessed via the Assess application on both the client's and practitioner's iPads. After the client is assessed, the completed assessment is stored in the Central portal and removed from the iPad. Reports can then be generated and the data can be exported for storage (Weiss et al., 2015). It should be noted that an internet connection is required for all steps outside of the actual assessment, which takes place via a Bluetooth connection between the devices.

Measuring Intelligence in Deaf and Hard-of-Hearing Students

Although the use of the digital version of the WISC-V is an issue for all clinical populations, for the deaf and hard-of-hearing population it poses a particular challenge. In general, intellectual abilities in individuals who are deaf or hard-of-hearing are difficult to measure due to several factors. Limited exposure to environmental sounds and spoken language, which is often the case for deaf and hard-of-hearing individuals, often impacts performance on verbal intelligence measures (J. P. Braden, 1985; Sullivan & M. Vernon, 1979).

In genthe, the history of intelligence tests is strongly tied to the production of speech and reasoning with language. The verbal portions of an intelligence test are a better measure of a deaf or hard-of-hearing individual's English proficiency than his or her underlying verbal reasoning abilities. This is significant to note due to the fact that the inappropriate use of psychological tests has been used to oppress deaf and hard-ofhearing individuals. Deaf and hard-of-hearing people were institutionalized based in part on their mental capabilities due to an underestimation of their abilities through the misuse of psychological tests (J. P. Braden, 1992). Early in the 20th century, research by Pintner showed that deaf and hard-of-hearing individuals scored lower on intelligence measures; thus, they were considered inferior to their hearing counterparts (Moores, 2001). As such, it is essential to ensure that deaf and hard-of-hearing individuals are being served ethically by any tests used to assess their intelligence. In a survey regarding test preferences for assessing deaf and hard-of-hearing people, practitioners serving this population in educational and clinical settings reported a strong preference for the Wechsler Performance Scales for assessing the intelligence in deaf and hard-of-hearing individuals (J. P. Braden, 1992). Practitioners also supported the use of the Chicago Non-Verbal Examination, Grace-Arthur Performance Scale, Hiskey-Nebraska Test of Learning Aptitude, Kaufman-Assessment Battery for Children, Leiter International Performance Scale, Ontario School Ability Examination, Snidjers-Oomen Nonverbal Test, WAIS-R, Wechsler-Bellevue Performance Scale, WISC- Performance Scale, WISC-R Performance Scale, Motor-Free Nonverbal Tests, Draw a Man/Person, Pinter Non-Language Test, and the Ravens Progressive Matrices. Given the strong preference for the Wechsler scales over other nonverbal measures, the equivalency between the paper and digital formats of the updated WISC-V must be established before it is used with this population.

Deaf and Hard-of-Hearing Normative Data

In the field of school psychology, there is an ongoing debate regarding the need of separate normative data to compare the performance of deaf and hard-of-hearing individuals on standardized testing (J. P. Braden, 1985). Deaf children vary in terms of the degree, onset, and etiology of hearing loss. Some children experience hearing loss prelingually, whereas others lose their hearing after having developed spoken language. The degree of hearing loss can vary from mild to profound and the etiology of the hearing loss may be a mix of sensorineural and/or conductive in nature. Furthermore, deaf and hard-of-hearing children also range in their exposure and access to spoken and visual language. Some deaf and hard-of-hearing individuals have access to spoken language through residual hearing or the use of technologies such as hearing aids and cochlear implants. Other deaf and hard-of-hearing individuals have access to visual language, such as American Sign Language or Cued Speech. A recent survey called the Regional and National Summary found that although 58.6% of children with hearing loss were identified as having no other conditions, the remaining children had at least one additional disability (Gallaudet Research Institute, 2013). All of these variables make creating norms for the deaf and hard-of-hearing population difficult.

Although some researchers have found that the use of special norms developed for deaf and hard-of-hearing students did not result in significantly different IQ scores (J. P. Braden, 1992), not all researchers support their use (Spencer & Marschark, 2010). One reason for this, as was concluded by J. P. Braden (1992), is that nonverbal tests yield substantially higher Iqs than verbal tests for deaf-and-hard-of-hearing people; however, poor administration practices account for lower performance in this population. Specific practices, such as test selection based on predicted performance, has confounded this research. Because deaf and hard-of-hearing students have average intelligence similar to their hearing peers on measures of fluid reasoning, one could make an argument for using the normative sample published with the test (Vernon, 1950).

Due to the many challenges of assessing these students, the NASP advocates for assessment using direct communication in the language and modality of a student (NASP, 2012). Furthermore, the NASP suggests that those assessing deaf and hard-ofhearing students should be aware of research in the field of deafness, specifically relating to the reliability and validity of psychological assessment instruments. Given the diversity in language proficiency, communication modality, and educational placement of students who are deaf or hard-of-hearing, they should be specifically included in the normative sample or validity studies of tests administered to this population.

Administration of Verbal Subtests

The use of verbal IQ scores with deaf and hard-of-hearing individuals is not supported in the literature and lacks evidence of validity (J. P. Braden, 1985; Sullivan & M. Vernon, 1979). Deaf and hard-of-hearing students often do not have the same incidental learning opportunities of their hearing peers. Since a significant number of deaf children are born to parents who do not use a visual language, these students are often not afforded the benefit of being exposed to the same amount of spoken language in their everyday environments (Conrad, 1979). Giving a verbal intelligence test to a deaf or hard-of-hearing student is instead a measure of his or her English proficiency, similar to that of an English language learner (NASP, 2012). A meta-analysis of 285 studies on the administration of intelligence tests on samples of deaf and hard-of-hearing students found a majority of practitioners used the Wechsler Performance Scales for assessing fluid reasoning in deaf and hard-of-hearing students (J. P. Braden, 1992). One limitation of this study was that a majority of the research was conducted in residential school settings; however, this is not uncommon given the low incidence of students who are deaf or hard-of-hearing. Additional studies have found that performance on other nonverbal intelligence tests for deaf and hard-of-hearing students were close to the mean performance of the tests' standardization samples (J. P. Braden, 1992). A more recent study found no mean composite score differences on the WISC-V between children with hearing differences who utilized spoken language and had assistive technology and a matched control group (Adams Costa, Day, & Raiford, 2016).

Summary

In summary, there is a body of research that supports the use of cognitive tests with deaf and hard-of-hearing children; however, there is no evidence of the appropriateness of using digital administration of the WISC-V with this population. Clinicians are currently using a psychological tool that has not been validated on the use of deaf and hard-of-hearing students. The current study was designed to provide data on the equivalency between the paper and digital administrations of the WISC-V with deaf and hard-of-hearing students. This study examined raw score equivalence between the standard and digital administrations of the WISC-V for students who were identified as deaf or hard-of-hearing. This research will add to the available literature to assist clinicians to make informed decisions when working with deaf and hard-of-hearing children.

Research Question and Hypothesis

Because there is documented equivalency of the paper-based and digital administrations of the WISC-V on hearing students (Daniel 2012), the purpose of this study was to determine whether raw scores are equivalent between paper-based and digital administration of the WISC-V for students who were identified as either deaf or hard-of-hearing. This study aimed to answer one research question: Is there equivalence of the digital-format and paper format of the WISC-V for students who are identified as deaf or hard-of-hearing?

Hypothesis. It was hypothesized that students would obtain the same number of correct items on the digital and paper administrations of the WISC-V.

CHAPTER 3: METHOD

Overview

This was a quantitative study in which a repeated measures experimental design was used. This study was a replication of the work published in the Q-interactive Technical Report 8 on the equivalency of the paper and digital versions of the WISC-V on hearing children (Daniel et al., 2014). The participants, deaf and hard-of-hearing students between the ages 6 through 16, took the Matrix Reasoning and Figure Weights subtests of the WISC-V twice, once in the traditional paper format and once in the digital format. The order of presentation of these two formats was counterbalanced across participants with half of the participants taking the paper version first and half taking the digital administration first followed by the alternative version.

Participants

The participants were recruited in the Mid-Atlantic region from 158 public schools and one school for the deaf. The participants who participated in the study were compensated with one free movie ticket, valued at less than 10 dollars. The participants who returned the parental consent forms were assigned to either Condition A (paper version first) or Condition B (digital version first) after being matched for gender, age, and degree of hearing loss. The students' ages and genders were collected via a demographic questionnaire completed by each child's parent or caregiver. This information went through a deidentification process to ensure student privacy. The complete demographic characteristics of all 22 participants is reported in Table 1.

Table 1

Demographic Characteristics of Students (N=22)

Administration Format		Paper First	Digital First
Gender	Male		
		5	4
	Female	6	7
Age (years)			
	7	1	0
	8	2	1
	9	3	2
	10	1	1
	11	0	2
	12	1	1
	13	1	1
	14	1	2
	15	0	1
	16	1	0
Degree of Hearing Loss/Deafness			
	Mild (26 to 40 dB)	2	1
	Moderate (41 to 55 dB)	4	5
	Severe/Profound (55-90 dB)	5	5
Equipment Used	Hearing Aids		
		7	7
	Cochlear Implant	3	4
	BAHA	1	0
	FM/DM	5	4
Mode(s) of Communication			
	Oral/Speech	6	8
	Cued Speech	3	2
	Sign Language	2	2
Communication with Child			
	Not Very Well	0	0
	Not Well	0	0
	Okay	0	0
	Good	3	1
	Completely	8	10
Home Spoken Language Used			
	English	8	7
	Non-English	3	4
Additional Disability			
	Yes	0	1
	No	11	10

Inclusion/exclusion criteria. Each participant was had a hearing loss of at least 20 decibels in the better ear as identified upon enrollment into an educational program for deaf and hard-of-hearing students, aged 6 years, 0 months through 16 years, 11 months, and identified as having an Individual Education Plan or 504 Plan. Since the WISC-V is given to the full range of heterogeneous population of deaf and hard-of-hearing students, participants in this study were not excluded based on the presence of additional disabilities. As this study used a repeated measures design methodology, each student acted as his or her own control.

Recruitment. The responsible adult on file for each student aged 6 years, 0 months through 16 years, 11 months who was identified as having a hearing loss, receiving special education services from a teacher of the deaf was sent a recruitment letter (Appendix A) and demographic questionnaire (Appendix B). The schools involved in the study identified the eligible students and controlled the distribution of the recruitment packets. Overall, 141 recruitment packets were mailed via the United States Postal Service and 208 recruitment packets were sent home in the backpacks of students. The packet included a preaddressed and posted envelope to return the parental consent and demographic questionnaire. There were a total of 24 packets returned, of which 22 contained parental consent and were included in the study. All of the parents completed the demographic questionnaires. The investigator was contacted by the parents of participants (2) who had questions about the study, ad several schools (5) contacted the investigator to ask questions. Some parents returned the recruitment material and declined testing (2) and some school principals declined to allow the study to be conducted in their buildings (4). A handful of principals did not respond to the study recruitment materials and their buildings were not used for the study (82). All participants that started the study completed and there was no attrition.

Sample size, power, and precision. Participants were selected based on geographic location that included a large suburban school system and a school for the deaf in the Mid-Atlantic region, where there are fewer than 400 students that meet the criteria for this study. An a-priori power analysis determined that the minimum sample size needed was 32 participants with an alpha = .05 and power = 0.80. Deafness is a low incidence population, and although this study used a convenience sample, the target sample size of this study was 50 participants but the study included 22 participants.

Measures and Materials

The measures that were used for this study were taken from the WISC-V. Given the lack of empirical support for the use of verbal measures to deaf and hard-of-hearing students, the use of the two primary Fluid Reasoning Index subtests were selected, Matrix Reasoning and Figure Weights. Due to the fact that deaf and hard-of-hearing students perform similarly to their hearing peers on this index, it is the most reliable of the five primary index scales on the WISC-V with this population.

On the Matrix Reasoning subtest, the participant viewed an incomplete matrix or series and selected the response option that completed the matrix or series. On the Figure Weights subtest, the participant, within a specified time limit, viewed a scale with missing weight(s) and selected the response option that kept the scale balanced. The total raw score of items answered correctly were summed to determine the participant's level of performance on each subtest in each condition. The average reliability across all age groups in the normative sample—which does not include children with a hearing loss—for the Matrix Reasoning subtest is .87 and the average reliability for the Figure Weights subtest is .94. A confirmatory factor analysis of the two subtests demonstrated .67 loading of Matrix Reasoning and .67 loading of Figure Weights on to fluid reasoning. This analysis supports the usage of the WISC-V as both a reliable and valid measure instrument.

Research Design

This study utilized a repeated measures design; each participant took the Figure Weights and Matrix Reasoning subtests in the paper and digital formats of the WISC-V. Given the diversity of the population being studied, this research design allowed for each participant to serve as their own control. This design was appropriate to use in this study, as the participants did not learn the solutions or new strategies for taking the subtests between administrations. That is to say that each student's performance was likely similar on both conditions. A repeated measures design was used in the original equality study (Daniel et al., 2014).

The retest equivalence was analyzed by calculating the mean difference between the first and second administrations. The mean value of difference should be the same regardless of sequence of test administration; however, if the mean difference scores between the two conditions differ by twice the size of the effect then there is a format effect. In order to detect an effect size of .2 (alpha = .05), a retest correlation of .8 and a sample of 22 cases (11 matched pairs) was used. Each of the format effects were obtained by computing the mean raw score changes in each format group. Next, the mean for the digital-first group was subtracted from the mean of the paper-first group and the results were divided by two.

Procedure

The research presented in this dissertation has been carried out according to the steps outlined to the Institutional Review Board of the Philadelphia College of Osteopathic Medicine. First, the investigator sent to each child's home a letter of recruitment, a parent consent form (Appendix D), a demographic questionnaire, and a postage paid envelope to return the completed documents to the investigator. The recruitment packet was sent to all 349 students who met the inclusion criteria. The parents or caregivers (N = 22) completed the consent and demographic survey through pencil-and-paper format and returned it via United States Postal Service. The survey consisted of 11 items and contained a mixture of Likert scale responses, open-ended responses, and multiple-choice responses. Participants for whom parental consent forms were returned were assigned to either Condition A, paper and then digital administration, or Condition B, digital and then paper administration, by matching groups for age, sex, and degree of hearing loss. The investigator coordinated with the participants' teachers to determine mutually agreed upon testing times. The investigator retrieved each student from his or her classroom and escorted the student to a room that was consistent with the testing environment as described in the Administration Manual of the WISC-V. The investigator obtained assent from the children (Appendix C).

The investigator thanked the participants for agreeing to help with this practice test, explained how long the tasks would take, shared that their participation was voluntary and that they could stop at any time, informed them that this was to help make tests better and they would not receive any grade for their performances, and asked whether they had any questions. The investigator read the following script: "Some of the things may be easy for you, but some may be hard. Just try your best." The investigator then administered either the paper or electronic version of the Matrix Reasoning and Figure Weights subtests using the script included in the test manual and followed all standardized testing procedures. The investigator thanked the participants for their help, reminded them that they would be taking the next part of the test again in approximately one week, and escorted them back to their classes.

The investigator again picked up the participants from their teachers approximately one week after the first administrations, repeated the above procedure, and administered the remaining format of either the paper or electronic version the Matrix Reasoning and Figure Weights subtests. The investigator thanked the participants for their help, gave them movie tickets for their participation, and walked them back to their classrooms. At the end of each testing session, the investigator transferred each participant's scores from the paper protocol or iPad to Statistical Package for Social Sciences (SPSS) by subject numbers to ensure the information was deidentified. The mean time between administrations for both conditions combined was 8.4 days. In summary, this study is a replication study of equivalency between the paperbased and digital administration of the WISC-V. This study was a quantitative repeated measures experimental design and was evaluated using a paired-samples *t* test. The participants were deaf and hard-of-hearing students between the ages 6 through 16, who took the Matrix Reasoning and Figure Weights subtests of the WISC-V twice, once in the traditional paper format and once in the digital format.

CHAPTER 4: RESULTS

The purpose of the present study was to investigate the equivalency of the paperbased and digital administrations of the WISC-V on deaf and hard-of-hearing students. The goal was to demonstrate equivalency between the paper and digital administrations of the WISC-V to ensure that the digital format is appropriate to use with this low incident population. The data gathered were analyzed using SPSS version 26.

The participants' raw score totals for both the paper condition and digital condition were analyzed using a paired-samples *t* test and using an effect size cutoff score of .2. Each participant's total raw score for both the Matrix Reasoning and Figure Weights subtests were recorded twice, once for paper format and once for digital format. All of the participants' mean raw score performances were compared for each condition to determine whether there were significant differences. The scores were then converted to an effect size by dividing by the standard deviation of each of the subtests by the population mean. The homogeneity of variance assumption was tested using a Pitman-Morgan test.

Data Entry, Scoring, and Survey

Data were collected from participants (N = 22) who were assessed using the WISC-V digital and paper versions. The number of correct items, raw scores, were totaled for both the Matrix Reasoning and Figure Weights subtests. Data were entered into SPSS to identify the total raw score (recorded as raw score) for Condition A and Condition B of the Matrix Reasoning and Figure Weights subtests, gender (1 = male, 2 = female), group condition (1 = paper first, 2 = digital first), hearing status of parents (1 = hearing, 2 = deaf/hard-of-hearing), age of hearing loss detection in months (reported as number of months), cause of hearing loss/deafness (1 = genetic, 2 = viral infection, 3 = medication, 4 = unknown), age early intervention started (recorded in number of months), degree of hearing loss/deafness (1 = mild [26 to 40 dB], 2 = moderate [41 to 55 dB], 3 = severe/profound [56 to 90 dB]), type of amplification used (1 =hearing aids, 2 = cochlear implants, 3 = bone anchored hearing aids [BAHA], 4 = FM/DM), additional disability (1 = no, 2 = yes), home spoken language (1 = English, 2= other), participant's preferred mode of communication (1 = oral/speech, 2 = cued speech, 3 = sign language), and parent ability to communicate with his or her child (1 = not very well, 2 = not well, 3 = okay, 4 = good, 5 = completely). Participants' raw score performances in Condition A, paper-based first, of Matrix Reasoning and Figure Weights were calculated and the descriptive statistics are provided in Table 2. Their raw score performances in Condition 2, digital first, of Matrix Reasoning and Figure Weights were calculated and the descriptive statistics are provided in Table 3.

Table 2 shows the mean performance of Condition A, paper-based administration first of the Matrix Reasoning and Figure Weights subtests of the WISC-V. The mean difference between the paper and digital administration of the Matrix Reasoning subtest was .46. The mean difference between the paper and digital administration of the Figure Weights subtest was .27. For this condition, both the Matrix Reasoning and Figure Weights subtests mean difference were less than .5 of a raw score point.

Table 2

Descriptive Statistics for WISC-V Subtests, Paper First

	Paper		**Q-interactive**	
Subtest	Mean	SD	Mean	SD
Matrix Reasoning	16.18	5.89	16.64	5.87
Figure Weights	18.73	5.86	19.0	6.05

Table 3 shows the mean performance of Condition B, digital-based administration first of the Matrix Reasoning and Figure Weights subtests of the WISC-V. The mean difference between the digital and paper administrations of the Matrix Reasoning subtest was .73. The mean difference between the digital and paper administration of the Figure Weights subtest was -.54. The difference between the digital and paper administration for both subtests was greater than .5 of a raw score point. For all conditions, the second administration was higher with the exception of Figure Weights in the digital-first condition. On this subtest, the mean performance of the participants between the first administration and second administration decreased. For this condition, both the Matrix Reasoning and Figure Weights subtests' mean differences were more than .5 of a raw score point.

Table 3

Descriptive Statistics for WISC-V Subtests, Digital First

Subtest	**Paper**		**Q-interactive**	
	Mean	**SD**	**Mean**	**SD**
Matrix Reasoning	18.55	5.53	17.82	5.52
Figure Weights	20.55	6.57	21.09	6.31

Table 4 shows the format effect and effect size for the paper-first condition for both the Matrix Reasoning and Figure Weights subtests of the WISC-V. The mean difference between the paper and digital administrations of the Matrix Reasoning subtest was .45. The mean difference divided by the standard deviation of the population of .52 results in a Cohen's D effect size of .87. The mean difference between the paper and digital administrations of the Figure Weights subtest was .27. The mean difference divided by the standard deviation of the population of 2.32 results in a Cohen’s D effect size of .11.

Table 4

WISC-V Effect Size, Paper First

Subtest	N	Mean Difference	SD	*t*	Effect Size
Matrix Reasoning	**11**	**-.45**	**.52**	**-2.88**	**.87**
Figure Weights	**11**	**-.27**	**2.32**	**-.38**	**.11**

Positive format effect indicates higher scores on paper administration.

Table 5 shows the format effect and effect size for the digital first condition for both the Matrix Reasoning and Figure Weights subtests of the WISC-V. The mean difference between the paper and digital administrations of the Matrix Reasoning subtest was .72. The mean difference divided by the standard deviation of the population of 1.0 results in a Cohen’s D effect size of .72. The mean difference between the paper and digital administrations of the Figure Weights subtest was .54. The mean difference divided by the standard deviation of the population of 1.86 results in a Cohen’s D effect size of .29.

Table 5

WISC-V Effect Size, Digital First

Subtest	N	Mean	SD	*t*	Effect Size
Matrix Reasoning	**11**	**-.72**	**1.009**	**-2.39**	**.72**
Figure Weights	**11**	**.54**	**1.86**	**.97**	**.29**

Positive format effect indicates higher scores on digital administration.

The format effects by subtest are shown in table 6. Each of the format effects was obtained by computing the mean first administration to second administration change score in each sequence group, then subtracting the mean for the digital-first group from the mean for the paper-first group, and finally dividing the results by 2. The results indicate that there is no format effect on the Figure Weights subtest; however, there is a small format effect on the Matrix Reasoning subtest for the deaf and hard-of-hearing participants in this study.

Table 6

WISC-V Format Effect by Subtest

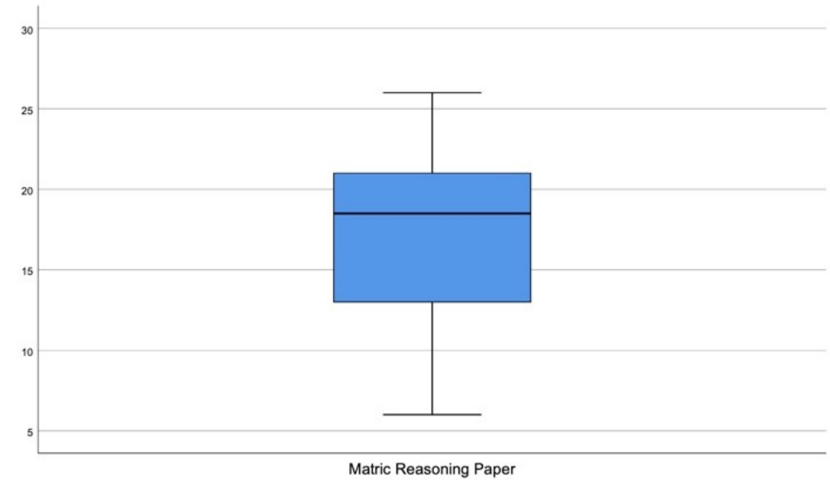
Subtest	**Format Effect**		***t***	**Effect Size**
	Mean	**SD**		
Matrix Reasoning	**-.95**	**-.72**	**-2.39**	**.31**
Figure Weights	**-.27**	**.54**	**.97**	**.09**

Positive format effect indicates higher scores on digital administration.

The results of this study were analyzed for possible outliers on either the Matrix Reasoning or Figure Weights subtests. The interquartile range indicated that all data points on both conditions for both subtexts did not include any outliers in the data. Figures 1 and 2 depict this graphically.

Figure 1

Matrix Reasoning Outliers


30
25
20
15
10
5
Matric Reasoning Paper

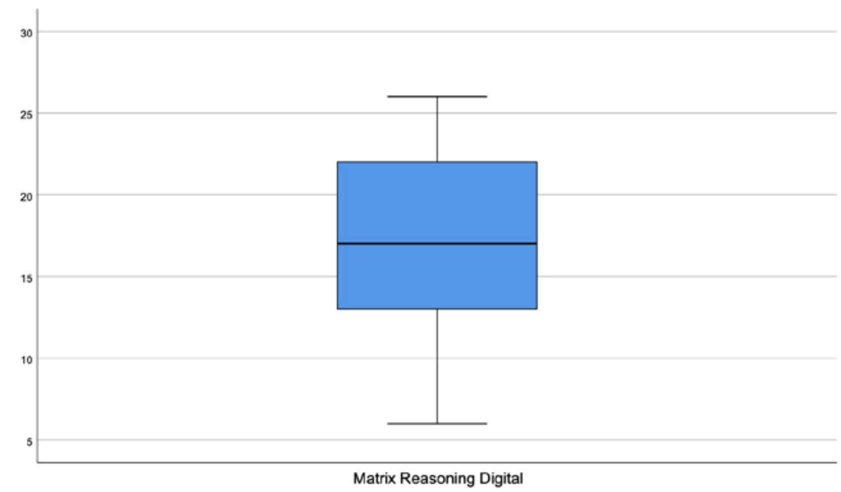
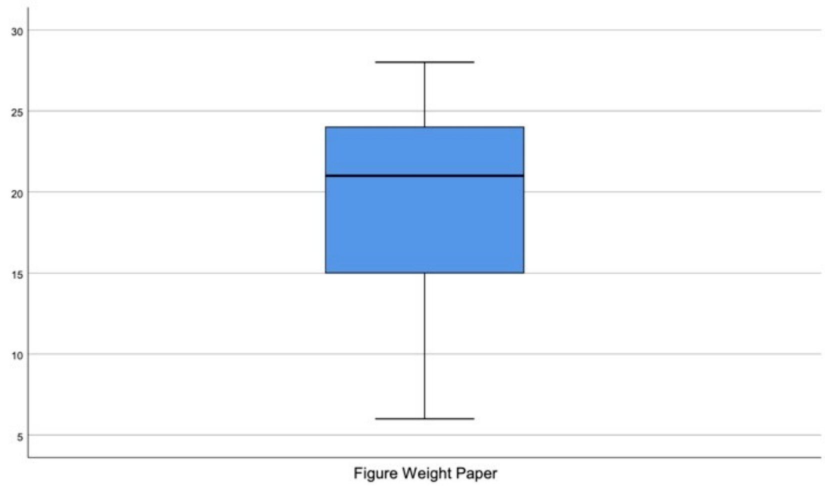
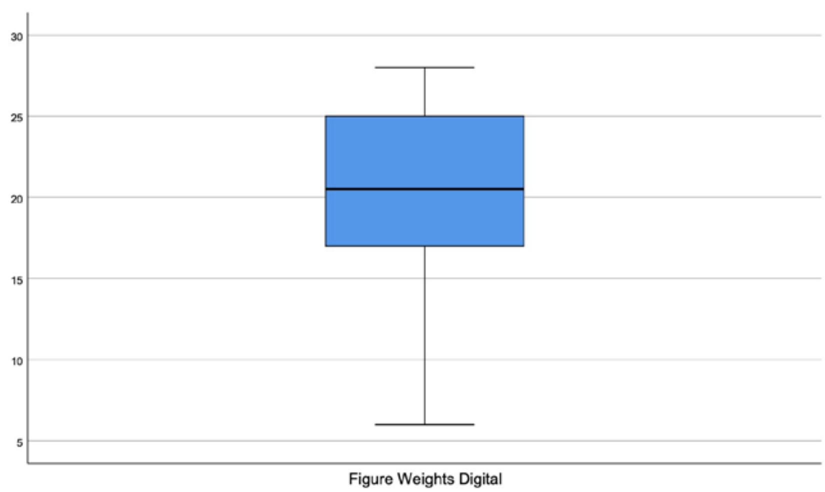

30
25
20
15
10
5
Matrix Reasoning Digital

Figure 2

Figure Weights Outliers





CHAPTER 5: DISCUSSION

There is documented equivalency of the paper-based and digital administrations of the WISC-V on hearing students (Daniel, 2012). As such, the purpose of this study was to determine whether raw scores are equivalent between paper-based and digital administrations of the WISC-V for students who are identified as either deaf or hard-ofhearing. Despite studies supporting the use of digital administration with special populations, no studies on the use of this format currently exist in the literature with the deaf or hard-of-hearing population. Since practitioners are currently using both the paper and digital formats of the WISC-V, it is essential to demonstrate that students would obtain similar scores regardless of format used. If equivalence is demonstrated, the norms, reliability, and validity information gathered for the paper format can be applied to the digital results for this population, even though deaf and hard-of-hearing students were not part of the normative sample. Given that deaf and hard-of-hearing children preform similar to hearing children on fluid reasoning measures (J. P. Braden 1992), the Figure Weights and Matrix Reasoning subtests are often the most reliable and valid subtests on the WISC-V to use with this population and were selected for this study. Although previous research has focused on the use of the WISC-V with deaf and hard-of-hearing students (Day et al., 2015), no study has shown equivalency of the digital and paper versions with this low incident population. The results do not fit with the previous equivalency studies that show children perform similarly on the paper and digital versions of the WISC-V (Daniel et al., 2014). Specifically, the effect size of the paper and digital administrations was less than .2 for the Figure Weights subtest. The effect size of the Matrix Reasoning subtest was .31, demonstrating a small effect size between scores on the two administration formats.

Interestingly, the mean score decreased on the Figure Weights subtest for the digital first group. This is unexpected, as performance would be predicted to remain the same or increase on the second administration. This decrease may be explained by fatigue or lack of interest in the paper-based material after taking the same subtest on the iPad. This may be useful for clinicians who have students who are near a cutoff score used to make a classification. Although there was no noticeable effect size between the formats on the Figure Weights subtest, it is important to note this unusual decline found in the data.

These results support that ability to interpret the results obtained from using the digital and paper formats the same way for the WISC-V Figure Weights subtest; however, the results of this study do not support the ability to interpret the results obtained from using the digital and paper formats the same way for the WISC-V Matrix Reasoning subtest with deaf and hard-of-hearing students. A small effect size was detected for the Matrix Reasoning subtest between the paper and digital formats. The results provide new data demonstrating deaf and hard-of-hearing children perform slightly better on the digital version of the Matrix Reasoning subtest compared to the paper version. These findings are similar with the available literature for the Figure Weights subtest and different for the Matrix Reasoning subtest when compared to previous work focusing on the equivalency between the paper and digital formats of the WISC-V (Daniel et al., 2014).

Impact of the Findings

The current study contributes to the existing literature on the use of the digital versus paper administration formats used with this low incident population. These results should be taken into account when considering format selection of the WISC-V when testing a child who is deaf or hard-of-hearing. The data contribute a clearer understanding on the ability of clinicians to use the norms, reliability, and validity information gathered for the paper format to the digital format of the Figure Weights subtest of the WISC-V. On the contrary, this study does not support the use of these resources for the digital administration of the Matrix Reasoning subtest of the WISC-V with deaf and hard-of-hearing students. The study provides new insight into the relationship between students who are deaf and hard-of-hearing and their interaction with digital and paper intelligence testing formats. Clinicians should use caution when choosing formats or interpreting the results of the Matrix Reasoning subtest of the WISC-V with deaf and hard-of-hearing students.

Limitations

Although the current study suggests a format effect between the paper and digital administrations of the Matrix Reasoning WISC-V with deaf and hard-of-hearing students, there are multiple factors impacting the ability to generalize these findings. The generalizability of the results is limited by the research being conducted using a convenience sample. The population of this study does not reflect the same level of geographic diversity as seen in the normative sample used by the test publisher. Additionally, socioeconomic status, race, and parental income levels were not collected as part of this study and may not reflect those provided in the test's normative sample, whereas they were part of the original equivalency study (Daniel et al., 2014). Furthermore, participants with other clinical conditions (e.g., students with ADHD, emotional disabilities, or learning disabilities) were not excluded from this study, possibly impacting the results. Lastly, due to the lack of available data on participants' motor skills and perceptual abilities, the results cannot confirm these possible factors confound the results of the study.

The methodological choices were constrained by the geographical location and access to participants. The size of the study sample was also a significant limitation, although not uncommon in low incident populations. The number of participants to meet the requirements of the power analysis was not satisfied. The study was designed to have at least 50 participants and only 22 out of 349 parents who received the recruitment packet provided consent. The low number of participants limits the ability to analyze the data by age, gender, hearing status of parents, cause of hearing loss, degree of hearing loss, or communication modality. A post hoc power analysis revealed the statistical power for this study was .60 for detecting a small effect size. The recruitment material and demographic questionnaire were only provided in English, likely impacting the ability for families from homes where languages other than English were spoken. This also limited the ability compare groups for differences based on communication modality and other factors.

Given that the research question for this study was focused on raw score equivalency between paper and digital formats for deaf and hard-of-hearing students, the results have not been analyzed based on age, sex, degree of hearing loss, parental hearing status, home language, communication modality, or presence of additional disabilities. The design of the current study does not account for the difference found between the paper and digital versions of the Matrix Reasoning subtest for deaf and hard-of-hearing students. Previous studies using randomly equivalent group design, non-random equivalent group design, and repeated measures design all found broad equivalency on nonclinical populations (Daniel, 2012; Daniel et al., 2014).

Future Directions

Further research is needed to confirm the use of the digital administrations of the WISC-V with deaf and hard-of-hearing students in order to demonstrate valid and reliable results compared to the paper version of the test. This study examined the differences between using the updated digital version of the WISC-V Matrix Reasoning and Figure

Weights subtests with a low incident population. Because the current findings on the equivalency with this population are mixed, more work is needed with this tool. A larger randomized study of students from a wider geographic area similar to the one used by the test publisher would be important to better represent students found in the United States. A larger study could also include recruitment material in several languages to include students from homes where languages other than English are spoken. Future investigations should explore whether equivalency between the paper and digital WISC-V for remaining subtests exists. It would also be important to ensure that in addition to the subtest raw scores, the composites and full scale scores demonstrate equivalency between the digital and paper formats. Although it was not a research question of this study, the data collected through the demographic questionnaire could be analyzed to determine the impact that gender, parental hearing status, degree and etiology of hearing loss, and modality have on performance on the paper and digital formats of the WISC-V. Lastly, future work could include other intelligence tests published on the QGlobal platform such as the WISC-V Spanish Edition, WAIS-IV, WPPSI-IV, and the Wechsler Memory Scale-Fourth Edition.

Although intelligence testing is a critical part of providing information about deaf and hard-of-hearing students to help provide educational services, there are other assessments that should also be explored. The variety of digital assessments that have been developed for online use in the past several years indicates the direction that the field of assessment is headed. Future research may wish to explore paper and digital equivalency of educational, language, executive functioning, and neuropsychological assessments that are also published on the Q-interactive system.

REFERENCES

Adams Costa, E. B., Day, L. A., & Raidford, S. E. (2016). *WISC-V special group study: Children with hearing differences who utilize spoken language and have assistive technology (WISC-V Technical Report #4).* Bloomington, MD: NCS, Pearson, Inc. Retrieved from https://images.pearsonclinical.com/images/Assets/WISC-V/WISC-V-Special-Population-Study-Assistive-Tech-Group.pdf

Braden, J. P. (1985). WISC-R deaf norms reconsidered. *Journal of School Psychology, 23*(4), 375-382. doi:10.1016/0022-4405(85)90050-0

Braden, J. P. (1992). Intellectual assessment of deaf and hard-of-hearing people: A quantitative and qualitative. *School Psychology Review, 21*(1), 82. Retrieved From https://www.tandfonline.com/doi/abs/10.1080/02796015.1992.12085597

Braden, J. S., DiMarino-Linnen, E., & Good, T. L. (2001). Schools, society, and school psychologists: History and future directions. *Journal of School Psychology, 39*(2), 203-219. doi:10.1016/S0022-4405(01)00056-5

Brue, A. W., & Wilmshurst, L. (2016). *Essentials of intellectual disability assessment and identification*. Hoboken, NJ: John Wiley & Sons, Inc.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytical studies.* New York, NY: Cambridge University Press.

Cattell, J. M. (1890). Mental tests and measurements. *Mind, 15*, 373-381. Retrieved from https://www.jstor.org/stable/2247264?seq=1

Conrad, R. (1979). *The deaf schoolchild: Language and cognitive function*. New York, NY: Harper Collins Publishers.

Cortiella, C., & Horowitz, S. H. (2014). *The state of learning disabilities: Facts, trends and emerging issues.* New York, NY: National Center for Learning Disabilities. Retrieved from https://www.researchgate.net/profile/Candace_Cortiella/publication/238792755_The_state_of_learning_disabilities/links/580e0cf508aebfb68a50953b.pdf

Daniel, M. H. (2012). *Equivalence of Q-interactive administered cognitive tasks: WAIS-IV (Q-interactive Technical Report 1).* Bloomington, MN: Pearson, Inc.

Daniel, M. H. (2013). *User survey on Q-interactive examinee behavior*. Bloomington, MN: Pearson, Inc.

Daniel, M. H., Wahlstrom, D., & Zhang, O. (2014). *Equivalence of Q-interactive and paper administrations of cognitive Tasks: WISC-V (Q-interactive Technical Report 8).* Bloomington, MN: Pearson, Inc.

Day, L., B., Adams Costa, E., & Raiford, S. (2015). *Testing children who are deaf or hard-of-hearing.* Technical report #2, NCS Pearson, Inc. Retrieved from https://www.pearsonclinical.com.au/files/8171450329565.pdf

Dearborn, G. V. N. (1927). theew of the abilities of man: Their nature and measurement. *The Journal of Abnormal and Social Psychology, 22*(3), 355-356.

DuBois, P. H. (1970). *A history of psychological testing.* Boston, MA: Allyn and Bacon.

Education for All Handicapped Children Act, U.S.C. 94-142 § 3500 (1975).

Elliott, C. D. (2007). *Differential ability scales*. San Antonio, TX: Harcourt Assessment.

Fiorello, C. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology, 14*(1), 2-12. doi:10.1080/09084280701280338.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (Vol. 84). Hoboken, NJ: John Wiley & Sons, Inc.

Galludet Research Institute. (2013). *Regional and national summary report of data from the 2011-12 Annual Survey of Deaf and Hard-of-Hearing Children and Youth.* Washington, DC: GRI, Gallaudet University.

Guilford, J.P. (1967). *The Nature of Human Intelligence*. New York, NY: McGraw-Hill.

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 41-68). New York, NY: The Guilford Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57,* 253-270.

Huertas, R. (2008). Between doctrine and clinical practice: Nosography and semiology in the work of Jean-Etienne-Dominique Esquirol (1772-1840). *History of Psychiatry, 19*, 123-40.

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Kaufman, A. S., & Kaufman, N. L. (2004). *KABC-II: Kaufman assessment battery for children*. Circle Pines, MN: American Guidance Service.

Mather, N., & Jaffe, L. E. (2016). *Woodcock-Johnson IV: Reports, recommendations, and strategies.* Hoboken, NJ: John Wiley & Sons, Inc.

In McCallum, R. S. (2017). Handbook of nonverbal assessment.

McGucken, W. J. (2008). *The Jesuits and education: The society's teaching principles and practice, especially in secondary education in the United States*. Eugene, OR: Wipf and Stock.

Moores, D. F. (2001). *Educating the deaf: Psychology, principles, and practices*. Boston, MA: Houghton Mifflin.

Morgan, C. L., & Steinman, C. C. (1943). An evaluation of a testing program in educational psychology. *Journal of Educational Psychology, 34*(8), 495-502.

National Association of the Deaf. (2016). Community and Culture–Frequently Asked Questions. Retrieved from https://www.nad.org/resources/american-sign-language/community-and-culture-frequently-asked-questions/

National Association of School Psychologists. (2012). *Sudents who are deaf or hard-of-hearing and their families [Position statement]*. Bethesda, MD: Author.

Otis, A. S. (1917). A criticism of the Yerkes-Bridges Point Scale, with alternative suggestions. *Journal of Educational Psychology*, *8*(3), 129-150.

Pick, L.H. (2013). Is it a sign of the times? Psychological and neuropsychological assessment of deaf people. In Peery, S. & Pick, L.H., *Practical considerations for the assessment of culturally and linguistically diverse people.* Symposium presen[te]d at the 121st Annual Convention of the American Psychological Association, Honolulu, HI.

Raiford, S. E., Drozdick, L., & Zhang, O. (2015). *Q-interactive special group studies: The WISC-V and children with autism spectrum disorder and accompanying language impairment or attention-deficit/hyperactivity disorder.* Bloomington, MD: Pearson, Inc.

Raiford, S. E., Holdnack, J., Drozdick, L., & Zhang, O. (2014). *Q-interactive special group studies: The WISC-V and children with intellectual giftedness and intellectual disability (Q-interactive Technical Report 9).* Bloomington, MD: Pearson, Inc.

Redvaldsen, D. (2017). Eugenics, socialists and the labour movement in Britain, 1865-1940. *Historical Research*, *90*(250), 764-787. doi:10.1111/1468-2281.12185

Roid, G. H. (2003). *Stanford-Binet intelligence scales* (p. 5). Itasca, IL: Riverside Publishing.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations*. San Diego, CA: Author.

Spearman, C. (1927). *The abilities of man.* London, UK: Macmillan.

Spencer, P. E., & Marschark, M. (2010). Evidence-based practice in educating deaf and hard-of-hearing students. Oxford: Oxford University Press.

Sternberg, R. J. (1988). What is intelligence?: Contemporary viewpoints on its nature and definition. Norwood N.J: Ablex.

Stern, W. (1914). The psychological methods of testing intelligence. Baltimore, MD: Warwick & York.

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by p racticing school psychologists: A national survey. *Journal of Psychoeducational Assessment*, *12*(4), 331-350.

Sullivan, P. M., & Vernon, M. (1979). Psychological assessment of hearing impaired children. *School Psychology Digest*, *8,* 271-290.

Terman, L. M. (1918). The use of intelligence tests in the army. *Psychological bulletin*, *15*(6), 177.

Terman, L. M. (1919). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the BinetSimon intelligence scale.* London, UK: Harrap.

Terman, L. M., & Merrill, M. A. (1973). *Stanford-Binet Intelligence Scale: 1972 norms edition.* Boston, MA: Houghton Mifflin.

Theriault, S. M. (2003). Patronage, the Pendleton Act, and the power of the people. *The Journal of Politics, 65*(1), 50-68.

Thurstone, L. L. (1975). *Primary mental abilities*. Chicago, Ill: University of Chicago Press.

U.S. Army. (1918). *Stanford-Binet Examination* [Database record]. Retrieved from PsycTESTS. doi:10.1037/t00012-000

U.S. Army, War Department. (1941). *General Classification Test* [Database record]. Retrieved from PsycTESTS. doi:10.1037/t00146-000

U.S. Department of Education. (2001). No child left behind act. Retrieved from https://www2.ed.gov/nclb/landin g.jhtml

Vernon, P. E. (1950). *The structure of human abilities.* London, UK: Methuen.

Wainer, H. (1988). The first four millennia of mental testing: From ancient China to the computer age. In *Proceedings of 29th Annual Military Testing Association Conference*, pp. 357-362.

Wechsler, D. (1939a). *Wechsler-Bellevue intelligence scale*. New York, NY: The Psychological Corporation.

Wechsler, D. (1939b). Wechsler-Bellevue Intelligence Scale, Forms I and II. New York, NY: The Psychological Corporation.

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

Wilkins.

Wechsler, D. (1949). *Wechsler intelligence scale for children.* New York, NY: The Psychological Corporation.

Wechsler, D. (1974). *Wechsler intelligence scale for children-revised*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler intelligence scale for children* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). Wechsler intelligence scale for children (4th ed.). San Antonio, TX: Pearson, Inc.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children: Technical and interpretive manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). Wechsler adult intelligence scale (4th ed.). San Antonio, TX: Pearson, Inc.

Wechsler, D. (2012). Wechsler preschool and primary scale of intelligence (4th ed.). Bloomington, MN: Pearson, Psychological Corporation.

Wechsler, D. (2014). Wechsler Intelligence Scale for Children (5th ed.). Bloomington, MN: Pearson, Inc.

Weiss, L. G., Saklofske, D. H., Holdnack, J. A., & Prifitera, A. (2015). *WISC-V assessment and interpretation: Scientist-practitioner perspectives*. San Diego, C: Academic Press.

Wundt, W., Creighton, J. E., & Titchener, E. B. (1894). Lecture first. In *Lectures on human and animal psychology*, Translated from German (2nd ed.) (pp. 1-11). New York, NY: Swan Sonnenschein & Co.

Yerkes, R. M. (1921). *Psychological examining in the United States Army*. Washington, DC: Washington Government Printing Office.

APPENDIX A: LETTER OF RECRUITMENT

THE EQUIVALENCY OF DIGITAL AND PAPER-BASED ADMINISTRATION OF THE
WECHSLER INTELLEGENCE SCALE FOR CHILDREN-FIFTH EDITION WITH DEAF AND HARD-OF-HEARING STUDENTS

My name is Kenneth Reimer and I am a student completing a doctorate degree at Philadelphia College of Osteopathic Medicine under the supervision of Dr. Katy Tresco. I am writing to invite your child to participate in my research study about taking an intelligence test on an iPad. I am completing this study to ensure that the deaf and hard of hearing students who take either the paper format or digital format of the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V) will obtain similar results.

In the study, your child will be given selected subtests from both the paper and digital formats of the test. Testing will take approximately two 20-minutes sessions that will occur one week apart. If the student agrees to participate, it will take about 40 minutes total of their time. During the test, students will look a series of pictures and select an answer from several possible choices. This test requires minimal language to complete and participation is completely voluntary. Any student with a hearing loss, between the ages of 6 years, 0 months and 16 years, 11 months, and attending the A school for the deaf and a large public school system in the Mid-Atlantic region is eligible to participate. In addition, parents or caregivers will complete a short demographic survey included with this letter. Students who choose to participate will get to use an iPad to complete one of the tests and receive a movie ticket for their participation. The study will take place at the student's school and done at a time that is convenient for the student and teachers to minimize disruption to instruction.

Your child's participation is completely voluntary. You can choose for them to be in the study or not. If you would like them to participate or have any questions about this study, please contact me at 301-965-0427.

Thank you very much.

Sincerely,

Kenneth Reimer

This study has been approved by the Philadelphia College of Osteopathic Medicine Institutional Review Board. For further information on this approval, please contact the Research Compliance Specialist at 215-871-6782.

APPENDIX B: DEMOGRAPHIC QUESTIONNAIRE

The purpose of this questionnaire is for you to provide some basic background information. Please complete the following questions.

1. Gender: ______ Male ______Female

2. Hearing status of parents: ______Deaf/Hard-of-hearing _______Hearing

3. Age hearing loss/deafness was detected: ______years _____months

4. Cause of hearing loss/deafness (if known)
_____Genetic
_____Viral infection
_____Medication
_____Other ____.

5. Age early intervention or IEP started: ______years _____months

6. Degree of hearing loss/deafness:
_____Mild (26 to 40 dB)
_____Moderate (41 to 55 dB)
_____Severe/profound (56-90 dB)

7. Does your child use:
______Hearing aids ______BAHA
______Cochlear Implant ______FM/DM

8. Please list any additional disabilities your child has: __________________________.
____.

9. Home language(s) used: ____.

10. Student's preferred mode(s) of communication:
_____Oral/Speech
_____Cued Speech
_____Sign Language

11. How well would you rate your ability to communicate with your child?

1…………………..2…..……………….3…..……………….4…..……………….5
Not very well Okay Completely

APPENDIX C: ASSENT FORM

Person in charge of the study: Mr. Ken Reimer Telephone Number: 202-821-2755

What is the study about?
Mr. Reimer wants to see if you will do as well taking a test on an iPad as you do taking a test on paper. If you want to be in the study, you will be asked to write your name on this form.

You do not have to be in the study. If you do not want to be in the study, that is OK, too. Don't put your name on the form if you don't want to be in the study.

What will happen to you if you are in the study?

You will be asked to do these things:
Take a test on an iPad.
Take the same test on paper.

How long will the study take?
The iPad test will take about 20 minutes. The paper test will also take about 20 minutes. So the study will take about 40 minutes total. If you say yes now and change your mind later, you can stop at any time. Just tell Mr. Reimer that you want to stop. Nobody will be angry with you if you say no now or later.

What if you have questions?
You can ask questions any time. You can ask now. You can ask later.

I understand what Mr. Ken Reimer has told me. I want to be in the study.

____
Child's Printed Name

____
Child's Signature

____

Date

APPENDIX D: INFORMED CONSENT FORM

TITLE OF STUDY

The Equivalency of Digital and Paper-Based Administration of the Wechsler Intelligence Scale for Children-Fifth Edition with Deaf and Hard-of-hearing Students.

TITLE OF STUDY IN LAY TERMS

The Reliability of the iPads to Administer the WISC-V Cognitive Test to Deaf and Hard-ofhearing Students.

PURPOSE

The purpose of this research is to find out This study is designed to demonstrate that the Qinteractive system will provide an equivalent score for individuals who are deaf or hard-of-hearing compared to the traditional paper-based test.

Your child is being asked to be in this research study because they have a hearing loss and are between the ages of 6 years, 0 months and 16 years, 11 months. If younger than 6 years or older than 17 years, cannot be in this study.

This study should take 40 minutes of your time.

INVESTIGATOR(S)

Principal Investigator: Katy Tresco, Ph.D.
Institution: Philadelphia College of Osteopathic Medicine
Department: Psychology
Address: 4170 City Avenue Philadelphia, PA 19131
Phone: 215-871-6630

Co-Investigator: Kenneth Reimer, Psy.S.
Institution: Philadelphia College of Osteopathic Medicine
Department: Psychology Address:
Phone:

Responsible (Student) Investigator: Kenneth Reimer

The test your child is being asked to volunteer for is part of a research project.

If you have questions about this research, you can call Dr. Katy Tresco at (215) 871-6630.

If you have any questions or problems during the study, you can ask Dr. Tresco, who will be available during the entire study. If you want to know more about Dr. Tresco's background, or the rights of research subjects, you can call the PCOM Research Compliance Specialist at (215) 871-6782.

DESCRIPTION OF THE PROCEDURES

If your child decides to be in this study, your child will be asked to take two subtests from the WISC-V cognitive test, once in a paper format and a second time on an iPad.

The study will take about 20 Minutes for each session . There will be 2 sessions over the course of 2 Weeks, for a total of 40 Minutes of your child's time.

POTENTIAL BENEFITS

Although your child may not benefit from being in this study. Other people in the future may benefit from what the researchers learn from the study.

RISKS AND DISCOMFORTS

Possible risks include invalidating the use of the WISC-V test again in the near future, requiring other psychological evaluations to choose a different test. The student may feel pressure to preform well on the test and that it might impacts their grades. To midigate these risks, your child will be told that this activity is to help make better tests and they will not be getting a grade or score.

ALTERNATIVES

The other choice is to not be in this study. Your child's participation is not required, there will be no consequences if they choose to not participate.

PAYMENT

Your child will be paid for being in this study. Your child will be provided with one free movie ticket for their participation, regardless of completion.

CONFIDENTIALITY

All information and records relating to your child's participation will be kept in a locked file. Only the researchers, members of the Institutional Review Board, and the U.S. Food and Drug Administration will be able to look at these records. If the results of this study are published, no names or other identifying information will be used.

REASONS YOUR CHILD MAY BE TAKEN OUT OF THE STUDY WITHOUT CONSENT

If health conditions occur that would make staying in the study possibly dangerous to your child, or if other conditions occur that would damage your child or your child's health, the researchers may take your child out of this study. You will be notified if your child is taken out of the study.

In addition, the entire study may be stopped if dangerous risks or side effects occur in other people.

NEW FINDINGS

If any new information develops that may affect your child's willingness to stay in this study, you will be told about it.

INJURY

If your child is injured as a result of this research study, your child will be provided with immediate necessary care.

However, your child will not be reimbursed for care or receive other payment. PCOM will not be responsible for any of your child's bills, including any routine care under this program or reimbursement for any side effects that may occur as a result of this program.

If you believe that your child has suffered injury or illness in the course of this research, you should notify the PCOM Research Compliance Specialist at (215) 871-6782. A review by a committee will be arranged to determine if the injury or illness is a result of your child being in this research. You should also contact the PCOM Research Compliance Specialist if you believe that your child has not been told enough about the risks, benefits, or other options, or that being pressured to stay in this study against your child's wishes.

VOLUNTARY PARTICIPATION

You and your child may refuse to be in this study. Your child voluntarily consents to be in this study with the understanding of the known possible effects or hazards that might occur during this study. Not all the possible effects of the study are known.

Your child may leave this study at any time.

If Your Child drops out of this study, there will be no penalty or loss of benefits to which entitled.

I have had adequate time to read this form and I understand its contents. **I have been given a copy for my personal records.**

I agree to allow my child to be in this research study.

Printed Name of Subject:____

Signature of Subject:____

Date: _____/_____/______ Time:______________AM/PM

Signature of Investigator or Designee_____________________________________
(circle one)

Date: ____/____/_________ Time:______________AM/PM