# INCREMENTAL VALIDITY AND THE WECHSLER INTELLIGENCE SCALES

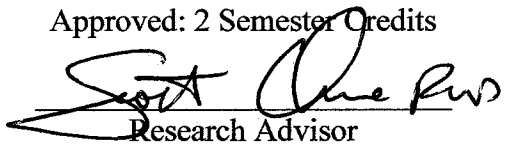
by

Rebecca S. Meyers

A Research Paper
Submitted in Partial Fulfillment of the
Requirements for the
Master of Science Degree
in

School Psychology

Approved: 2 Semester Credits



Research Advisor

The Graduate School

University of Wisconsin-Stout

August, 2004

The Graduate School

University of Wisconsin Stout

Menomonie, WI 54751

# ABSTRACT

Meyers (Writer) (Last Name) Rebecca (First Name) S. (Middle Initial)

Incremental validity and the Wechsler Intelligence Scales
(Title)

School Psychology (Graduate Program) Scott Orme, Ph.D. (Research Advisor) August/2004 (Month/Year) 24 (# of Pages)

American Psychological Association, 5th edition
(Name of Style Manual Used in this Study)

Throughout history, there have been several theories regarding the concept of human intelligence. Often, these theories play a major role in the development of assessment tools used to help identify children who are receive special education services. Currently, there are several new tests that are being developed that are used to make special education decisions. It is essential to examine how new tests or subtests of an assessment battery provide additional and useful information about special education decisions and services. A review of literature will be examined to determine what valuable information the new subtests and theoretical framework of the Wechsler Intelligence for Children-4th Edition (WISC-IV) provide. Additionally, attention will focus on determining if incremental validity is a validating approach.

TABLE OF CONTENTS

# CHAPTER I: INTRODUCTION

The concept of intelligence and intelligence assessment has a long and growing history. IQ tests are deemed beneficial when they can provide information about how children learn (Esters, Ittenback, & Han, 1997). In a given year, schools in the United States typically have close to 10% of all students receiving services in special education and IQ tests often play a major role in the entire special education services process. Results from IQ tests often provide the framework for reports that are often a great resource for determining service related questions and decisions. Currently, laws seem to dominate and control school psychology practice in schools in two major areas. First, legal requirements are typically the funding source for school psychology and secondly, eligibility for special education services is the main focus for school administrators (Esters, Ittenback, & Han, 1997). Consequently, it is essential to review and examine past and current tests to determine how they differ and how current tests have improved. One of the leading tests used in schools has been the Wechsler Intelligence Scale for Children (WISC) in various revisions. The latest version (WISC-IV) was released in fall 2003. In the new WISC-IV several changes from old versions have been made, in particular, several new subtests have been added (Wechsler, 2003). As such, it will be necessary to determine specifically how the new WISC-IV has improved with the addition of several new subtests.

Currently, theories of intelligence play a major role in the development of new assessment tools. However, in the past, this was not always the case. Specifically, the Measuring Scale of Intelligence focused on creating items that would differentiate children that would benefit from traditional education. The Wechsler Intelligence Scale

for Children focused on problem solving ability. Wechsler proposed that children demonstrate their intelligence verbally and motorically. The 1905 Binet-Simon scale and the first WISC, developed by David Wechsler (1939), were created by clinicians, and research does not support any formal theory supporting or underlying either test, however, today academics, developers, and practitioners believe that a theory of framework is essential for creating an assessment tool. The Binet Scales and WISC could both provide information about observable abilities, but lacked in providing information regarding underlying abilities (Esters, Ittenback, & Han, 1997).

In this chapter attention will be give to the background of early assessment tools and current theories surrounding the concept of intelligence. Additionally, the link between current theory and how it is part of assessment will be discussed. Specifically, the Woodcock Johnson III and the Wechsler Intelligence Scale for Children-Fourth Edition will be examined to determine now current theory plays a role in these two tests. Furthermore, particular attention will be covered regarding the additional subtest in the new WISC and determine if these tests are useful for diagnosing and implementing strategies and interventions in schools.

*Early Theoretical Development*

Typically, factor analytic theorists can be categorized into two separate groups: theorists who prefer a general-factor (g) theory of intelligence and theorists who promote a multiple-factor theory. Charles E. Spearman promoted factor analysis, while both Edward L. Thorndike and Lousis L. Thurstone advocated that intelligence is made up of several independent factors (Sattler, 2001).

Spearman is known to be an early advocate of factor analytic approach to the concept of intelligence (Sattler, 2001). Spearman suggested a theory of intelligence that involves two factors that help provide an explanation of patterns of correlations in group tests of intelligence. Spearman's theory proposed that a factor (*g*) in addition to a more specific factor (*s*) per test explain for one's performance on a given intelligence assessment. Spearman's explanation of *g* involved the *g* factor known as a general mental energy. The *g* factor is involved in complex operations that involve ability, quickness, intensity, and degree of intellectual productivity (Sattler, 2001).

Spearman also deems the *g* factor to be an index of general mental ability or intelligence and it is considered to be the "inventive" part of mental ability instead of the "reproductive" aspect of mental ability (Sattler, 2001). Spearman would consider tests with high *g* loadings to consist of reasoning, comprehension, matrix tasks, generalizations, verbal analogies, math problems, paragraph comprehension, and perceptual analogies (Sattler, 2001).

Thorndike proposed a multifactor theory that involved interrelated, but separate intellectual abilities (Sattler, 2001). Thorndike suggested that mental activities have certain similar attributes and together they create clusters. He identified three clusters: social intelligence, concrete intelligence, and abstract intelligence. More specifically, social intelligence involves interacting with other individuals, concrete intelligence involves interacting with objects, and abstract intelligence involves working with verbal and mathematical characters. Finally, it is noteworthy to include that Thorndike's perceptions involve a theoretical viewpoint, instead of a statistical method(s) (Sattler, 2001).

Unlike Spearman, Thurstone did not agree that intelligence was made up of a single element (Sattler, 2001). Instead, Thurstone proposed, "human intelligence possesses a certain systematic organization, with a structure that we can infer from a statistical analysis of the patterns of intercorrelations found in a group of tests" (Sattler, 2001, p. 138). Thurstone used a factor analysis method that could examine several factors at the same time to determine primary mental abilities. He consequently identified verbal, perceptual speed, inductive reasoning, number, rote memory, deductive reasoning, word fluency, and space or visualization as primary mental abilities. Additionally, he deemed that one could separate intelligence into those primary mental abilities with each being equally weighted in significance. Despite agreeing with *g* as an important aspect of mental functioning, later research revealed that these primary abilities correlated moderately, consequently leading Thurstone to hypothesize the contribution of a second-order ability associated with *g* (Sattler, 2001).

*Early Assessment Methods*

Sir Francis Galton's statistical studies of mental processes shaped much of what would occur in the 20th century for psychology (Sattler, 2001). Galton's contribution to the field of psychology included the development of regression to the mean and correlation. These developments lead to the study of intelligence and the study of relationships between the intelligence scores of parents and their kids. Galton published *Hereditary Genius* (1869), which discussed a statistical reason for mental attributes and proposed the amount of "geniuses" that could be projected in a specific sample of people. Later, Galton wrote and published *Inquiries into Human Faculty* (1883) that discussed his perceptions on human faculties and the problems surrounding the ability to measure

mental attributes. In 1884 Galton created a psychometric laboratory housed at the International Health Exhibition, later established at University of College, London. Galton measured physical and mental capacities because he proposed that high intelligence is associated with excellent sensory discrimination abilities. This assumption paved the road to tests of sensory discrimination and motor coordination to aid in the study of mental performance. However, these tests proved to be invalid measures and thus limited his involvement on the measurement of intelligence (Sattler, 2001).

At the end of the 19th century, Alfred Binet, Victor Henri, and Theodore Simon proposed the idea that one should focus on higher mental processes rather than simple sensory functions for determining intelligence (Sattler, 2001). Consequently in 1905, Binet and Simon created an intelligence scale that would identify children who would not succeed in a traditional education environment and were considered to have mental retardation. In contrast to previous assessment tools, this scale recognized age-based cognitive development. Additionally, it provided degrees of mental retardation and became the model for future scales that examined mental ability. In 1908 and 1911 Binet and Simon revised their scale to accurately decipher whether a child was performing at the average level of children of the same age or not. Lewis M. Terman and his colleagues at Stanford University made a revision and translation into English, but it was still concentrated on distinguishing intellectual deficiency in children. With the help of H.G. Childs, Terman published the Stanford Revision and Extension of the Binet-Simon scale that included standardization. Additionally, Terman embraced Louis William Stern's idea of mental quotient, which involves dividing the mental age by chronological age. Terman and his colleagues later renamed this ratio the intelligence quotient in the

1916 revision of the scale. In 1937 and 1960, Lewis Terman and Maud Merrill revised the Stanford-Binet. In 1972, updated norms were published and Deviation Quotients were included instead of the ratio IQ. Robert Thorndike, Elizabeth Hagen and Jerome M. Sattler in 1986 created a point-scale version of the assessment called the Stanford Binet Intelligence Scale: Fourth Edition (Sattler, 2001). In 2003 the new Stanford Binet Intelligence Scale-Fifth Edition (SB:V) was developed (Roid, 2003). The new SB:V includes both routing sublets (subtests) that are in the point-scale format and is based on the modern item response theory (IRT) for a psychometric foundation. The SB:V incorporated "new rules of measurement" that included routing techniques that enhanced the accuracy of measurement by accommodating the level of item difficulty to match the level of the examinee's current level of functioning (Roid, 2003).

David Wechsler also took notice to the idea of a point scale (Sattler, 2001). Consequently, in the 1930s Wechsler examined standardized tests and ultimately chose 11 subtests to form the first Wechsler intelligence test, the Wechsler-Bellevue Intelligence Scale. This scale was the predecessor to the Wechsler Intelligence Scale for Children-Third Edition, Wechsler Preschool and Primary Scale of Intelligence-Revised, and the Wechsler Adult Intelligence Scale-Third Edition. Wechsler thought of intelligence to be component of the larger concept of personality. As such, Wechsler included subtests that were intended to provide psychiatric diagnosis. Additionally, the IQ score was deemed an index of general mental ability (Sattler, 2001).

*More Recent Developments in Cognitive Assessment*

During the 1950s special education started to expand, and assessment shifted more to measuring cognitive functioning. Simultaneously, Cattell developed a theory of intelligence that involved fluid intelligence and crystallized intelligence (Sattler, 2001).

Fluid intelligence is considered nonverbal and cultural free. Some tasks that involve fluid intelligence include: figure classifications, figure analyses, number and letter series, matrices, and paired associations (Sattler, 2001). Typically, fluid intelligence develops well into adolescence and then gradually declines due to deterioration of physical structures. The ability to perform tasks quickly and store and manipulate stimuli (working memory) are both components of fluid intelligence (Sattler, 2001).

Crystallized intelligence involves skills that are dependent on culture. There are several components that are included in crystallized intelligence: vocabulary, general information, abstract word analogies, and mechanics of language (Sattler, 2001). Additionally, there are several informal and formal aspects that influence and impact one's crystallized intelligence. Additionally, there are some tasks that are components of both fluid and crystallized intelligence. Specifically, arithmetic reasoning, inductive verbal reasoning, and syllogistic reasoning all contribute to both of these structures of intelligence (Sattler, 2001).

Horn disagreed with the notion of general intelligence, indicating that research lacks support of a unitary theory (Sattler, 2001). Alternatively, he proposed that intellectual ability is comprised of numerous specific functions that have a genetic link and result in different paths for development over the course of one's life. For example,

fluid ability and visually thinking often decrease as one becomes older, while crystallized ability and long-term acquisition and retrieval remain stable. As time elapsed, Horn added to Cattell's theory and included visual perception, short-term memory, long term storage and retrieval, speed of processing, auditory processing ability, quantitative ability, and reading and writing ability factors (Sattler, 2001).

John B. Carroll developed a three-stratum factor analytic theory that was focused on cognitive abilities (Sattler, 2001). He analyzed 465 research studies and determined that there are several specific individual differences in cognitive ability. Carroll proposed that these differences resulted in three groups or strata. The first stratum is referred to as the narrow stratum and includes 65 narrow abilities that involve one's level of expertise in several cognitive areas. For example, general sequential reasoning, reading comprehension, memory span, visualization, speech sound discrimination, creativity, numerical facility, and reaction time are all measures of narrow abilities. The second stratum involves fluid intelligence, crystallized intelligence, memory and learning, broad visual perception, broad auditory perception, broad retrieval capacity, broad cognitive speediness, and processing speed. The third stratum involves a general factor, or *g* (Sattler, 2001).

As a result of Carroll's analyses, his findings were combined with the theoretical information of Cattell and Horn as the Cattell-Horn-Carroll or CHC Theory of Intelligence (Horn & Noll, 1997). This also is often called the Gf-Gc Theory. This model is currently the most widely used and accepted amongst test developers and more and more practitioners are beginning to accept it (Horn & Noll, 1997).

*Current Status of Cognitive Tests*

The latest versions of the major instruments are designed to provide clinical utility with theoretical support following CHC model (Horn & Noll, 1997). The test that most matches the various hierarchical structure is the Woodcock Johnson. It is comprised of 20 subtests measuring various skills within the academically relevant domains of CHC theory. The recent Stanford Binet-Fifth Edition (2003) is also tailored along its broad scales to match this theory. Earliest versions of the WISC defined intellectual functioning along a Verbal/Performance split. The new WISC-IV is a departure from that approach in that it tries to more closely fit within the CHC theory (Horn & Noll, 1997).

*Statement of the Purpose*

The purpose of this literature review is to review the WISC-IV and incremental validity as a validating approach. The following research questions guided this review:

1. What is the rationale of including new subtests into the WISC-IV battery?
2. How are these additional subtests providing novel and valuable information that will help school professionals make decisions regarding special education services and decisions?
3. How have past studies regarded the value of the WISC in making special education decisions?

# CHAPTER II: LITERATURE REVIEW

This chapter will focus on the new Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV). Specifically, it will be essential to determine the rationale behind incorporating new subtests into the battery. Additionally, it is important to determine if these subtests are providing new and useful information to help make decisions regarding special education services for students. Lastly, existing literature will be reviewed to determine how past studies have evaluated the usefulness of the WISC to make special education decisions.

## *The WISC-IV*

The Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) is an inclusive assessment tool that help assess intelligence in children 6 years 0 months to 16 years 11 months (Williams, Weiss & Rolfhus, 2003). The WISC-IV helps identify certain diagnoses and encourages implementation of useful interventions for children. The new WISC-IV also includes a new neurocognitive model of information processing which replaces the previous Verbal IQ (VIQ)/Performance IQ (PIQ) dichotomy. Specifically it provides composite scores that are represented in certain domains: Verbal Comprehension Index, Perceptual Reasoning Index, Working Memory Index, and Processing Speed Index in addition to including a Full Scale IQ representation of intellectual ability. Lastly, the authors also included five new subtests to this theoretical model (Williams, Weiss, & Rolfhus, 2003).

The authors of the WISC-IV determined that before introducing a new test, several concerns must be addressed (Williams, Weiss, & Rolfhus, 2003). Over the

course of five years, the new WISC was given to several thousands of children and the results of these administrations were reviewed for cultural, socioeconomic, and/or regional bias and concerns. The Psychological Corporation and outside experts deemed only the best items to be refined to include in the most recent rendition of the WISC (Williams, Weiss, & Rolfhus, 2003). .

The WISC-III and the WISC-IV also differ in their focus for the Full Scale IQ (FSIQ). For example, the new WISC-IV focuses more on fluid reasoning, working memory, and processing speed instead of primarily crystallized knowledge (Williams, Weiss, & Rolfhus, 2003). In addition, the authors added three new core subtests and two supplemental tests to the WISC-IV. Picture Concepts, Letter-Number Sequencing, Matrix Reasoning, and Word Reasoning are subtests that were modified from additional Wechsler Intelligence Scales and will be further described in the following sections. Cancellation is a completely new subtest that was developed (Williams, Weiss, & Rolfus, 2003).

*Subtests of the WISC-IV*

Picture Concepts, is a new subtest that is part of the core Perceptual Reasoning Index scale. This task involves the child identifying a picture that shares a similar characteristic with a group of pictures. This subtest has 28 questions that are intended to examine one's fluid reasoning and abstract categorical reasoning. The beginning items focus on one's ability to reason based on the item's color, shape and appearance, while advanced items examine one's ability to determine the function of an object (Williams, Weiss, & Rolfhus, 2003).

Letter-Number Sequencing is a new subtest that is part of the core Working Memory scale. This task consists of ten items of three trials each that include a sequence of numbers and letters the examinee is asked to recall in a certain order; numbers from smallest to largest, and letters in alphabetical order. Letter-Number Sequencing measures one's sequencing, ability to manipulate stimuli, attention and focus, short-term auditory memory, visual-spatial imaging, and ability to complete tasks quickly (Williams, Weiss & Rolfhus, 2003).

Matrix reasoning is a core Perceptual Reasoning subtest that involves the examinee choosing the missing portion of a matrix from five choices. Matrix reasoning measures one's fluid reasoning and general intellectual ability (Williams, Weiss & Rolfhus, 2003).

Word Reasoning is a new supplemental subtest that is part of the Verbal Comprehension subtest. Word Reasoning involves the examine identifying a general theme that is verbally described in a set of clues. Word Reasoning tasks may help determine how one comprehends verbally, one's reasoning skills, the ability to combine and synthesize information, and produce alternative concepts (Williams, Weiss & Rolfhus, 2003).

Cancellation is a supplemental subtest that is part of the Processing Speed subtest. The examinee is asked to identify certain items in either a structured arrangement of items or unstructured arrangement within a certain time frame. Cancellation determines how fast one can identify visually similar objects (Williams, Weiss & Rolfhus, 2003).

*Theoretical Foundation of the WISC-IV*

The WISC-IV concentrated on improving the theoretical foundations of the test, improving clinical applications, improving psychometric properties, improving user friendliness, and preserve stability and familiarity (J. A. Hanson, personal communication October 10, 2003). For the purpose of this review, it will be important to focus on how the new WISC-IV improved clinical utility. The developers of the WISC-IV concentrated on evaluating scores and performance for these groups: intellectually gifted; children with mild or moderate mental retardation; children with reading, written expression, and mathematical disorders; children with learning disabilities and attention-deficit/hyperactivity disorder; children with expressive and/or receptive language disorders; children with traumatic brain injury; children with Asperger's disorder; and children with motor impairment (Williams, Weiss, & Rolfhus, 2003).

*Interpretation of the WISC-IV*

Interpretation of the WISC generally involves the use of profile analysis to look at one's strengths and weaknesses in cognitive abilities (Sattler & Dumont, 2004). The Full Scale IQ (FSIQ) is an important measure of general intellectual ability, but lacks the ability to provide information about the fundamental abilities from which it is derived. Profile analysis strives to go beyond evaluating one's overall functioning and determine unique ability patterns. Additionally, information received from ability patterns can help teachers develop successful strategies and implement interventions. The main goal of profile analysis involves producing hypotheses regarding the individual's abilities (i.e., strengths and weaknesses). One can evaluate profiles in two different ways. The first comparison involves examining the individual's scores to the normative sample, i.e. an

inter-individual comparison. The second comparison involves comparing the individual's scores to their performance, i.e. an intra-individual comparison. Both comparisons use scaled scores from the normative sample (Sattler & Dumont, 2004).

The most straightforward method to determine subtest profile analysis is to examine the scores based on the normative sample or inter-individual comparison (Sattler & Dumont, 2004). One can evaluate and describe subtest scaled scores by using either a three-category or a more comprehensive five-category approach. In both methods, scaled scores of 13 to 19 always specify strength and scores of 8 to 12 specify average ability, and scores of 7 or below indicate that the individual does not strengths relative to peers of their same age. However, it is important to note that the individual may have strengths and weaknesses that were not assessed by the test. Percentile ranks should also be included and they help provide a more accurate indicator of the individual's level of functioning. The other method that examines profile analysis involves using terminology geared toward the individual's performance. For example, using descriptions that involve one's own relative performance to their own level of functioning is preferred (Sattler & Dumont, 2004).

When the Wechsler scales and similar tests were developed, many professionals believed that profile analysis would aid in clinical utility and more specifically help determine diagnoses (Sattler & Dumont, 2004). However, results did not provide clinical results they were counting on. Many psychologists concluded that "Thus, profile analysis with the WISC-IV cannot be used to arrive at a diagnostic label" (Sattler & Dumont, 2004, p. 114). They concluded that Index scores are more reliable than subtest scaled scores and subtests and individual Composites measure distinct cognitive processes.

Fortunately, valuable information such as one's cognitive strengths and weaknesses can be determined and provide insight about the individual's functioning (Sattler & Dumont, 2004). Consequently, it would be beneficial to research the new WISC-IV and determine how these new subtests can provide insight about incremental validity and specifically gain information regarding one's strengths and weaknesses.

*Concept of Incremental Validity*

There have been a few past research studies that have examined incremental validity in the WISC-III to determine how it may provide insight in academic performance. According to Hunsley and Meyer (2003), although researching incremental validity appears uncommon, it is a very important measure to use when a new test is developed or a new subtest is included to a previously existing scale. Specifically, it is essential to provide information that the additional scale included information that was not already available or less adequately obtained. The lack of incremental validity research would further escalate the continual production of reconfigured items or variables (Hunsley & Meyer, 2003). According to Hunsley and Meyer several analysts have suggested an increased focus on incremental validity in interviews, observations, and psychological tests. However, there continues to be a lack of research that systematically examines incremental validity (Hunsley & Meyer, 2003 The following studies provide examples of how researchers have examined incremental validity in the WISC-III.

*Incremental Validity of the WISC-III*

According to Watkins and Glutting (2000) many researchers believe that despite the notion that profiles from cognitive assessments do not provide correct information to

make diagnoses for childhood psychopathology, researchers still use subtest profiles to speculate children's ipsative strengths and weaknesses from cognitive assessment tools. Many researchers conclude that subtest profiles do not significantly predict achievement, but Watkins and Glutting (2000) believe that this is due to constraints regarding participants and instruments. Their rationale suggested that there have been few studies that have looked at incremental validity of the WISC-R and the WISC-III. Watkins and Glutting (2000) also thought it was necessary to examine the WISC-III because it included many revisions of materials and administration processes. Additionally, they noted that past studies include applied criterion achievement assessments that are beneficial for screening purposes, but may be insufficient of determining academic achievement for determining incremental validity. Lastly, past research studies have only used small samples of individuals in their assessment to determine eligibility for special education services.

According to Watkins and Glutting (2000) other researchers have suggested that there are biases in special education referrals and potentially subtest profiles may be helpful for predicting academic achievement, however sampling error remains a concern. Also, the shape or ups and downs of one's subtest profile may provide insight in students, who are considered exceptional, but not with students who are considered nonexceptional or vice versa. As a result, Watkins and Glutting (2000) conducted a study that examined the incremental validity of the WISC-III profile elevation, scatter, and shape for simultaneously predicting broad measures of reading and math performance.

Their study consisted of two sample groups: nonexceptional sample and exceptional sample. The WIAT was give to the nonexceptional sample, the WJ-R was

given to the exception sample to determine academic achievement and the WISC-III was given to both samples. Watkins and Glutting conducted several regression analyses that were intended to predict numerous achievement variables. Their study included analysis of profile elevation, followed by profile scatter, and then profile shape was entered last (Watkins & Glutting, 2000).

Results indicated that profile elevation is predictive of academic achievement in both exceptional and nonexceptional students (Watkins & Glutting, 2000). However, results also indicated that subtest profile scatter did not provide a prediction of achievement despite students demonstrating clinical scatter on a minimum of one subtest. Additionally, profile shape accounted for an additional only 5%-8% of the variance in achievement measures and thus increased prediction minimally. Watkins and Glutting concluded that incorporating WISC-III subtest scatter and shape is not indicative of academic performance (Watkins & Glutting, 2000).

In another article, researchers Glutting, Youngstrom, Ward, Ward, and Hale (1997) also wanted to examine incremental validity in the WISC-III. Specifically, they wanted to determine if the four factor scores from the WISC-III predict achievement on the Wechsler Individual Achievement Test (WIAT). Glutting and Youngstrom included two samples of children in their study: a nonreferred sample and referred sample. The nonreferred sample was comprised of children who had been administered the WISC-III, WIAT, and the Guide to the Assessment of Test Session Behavior. The referred sample consisted of individuals who had been referred for psychoeducational assessments from their school district (Glutting, et al., 1997).

The research team assessed the contributions of the four factor scores to achievement criteria through numerous hierarchical regressions. The dependent variable consisted of the Reading, Mathematics, Writing, and Language scores of the WIAT in one regression analysis. Glutting et al. (1997) completed a hierarchical analysis comparing the FSIQ to the other broad scales of the WISC-III to see which provided the best predictor of academic success.

Results from the study indicated that verbal and performance IQs provide only a minimal prediction of any achievement criteria in this study. In general, scores from the PIQ or VIQ did not contribute any additional information that was not already provided by one's FSIQ for prediction in reading, math, writing, or language performance on the WIAT. Additionally, results indicated that the verbal comprehension index (VCI) did not provide substantial contribution in children that were not referred and provides only a minimal increment in children that were actually referred from assessment (Glutting et al., 1997).

The Freedom from Distractibility Index (FDI) showed a small, but significant relationship to reading, math, and writing from the WIAT. Even though the FDI illustrated the largest part correlations compared to other ability scores, it only accounted for 2% of the variance in achievement. Glutting et al. (1997) determined that the FDI alone does not provide sufficient additional data regarding one's achievement to warrant the extra time exploring it.

The Processing Speed Index did not provide additional information about predicting one's performance in reading or math. However, it did provide some evidence

of predicting writing or language, but these the researchers deemed it too minimal to provide clinical utility (Glutting et al., 1997).

Overall, Glutting and colleagues deemed the FSIQ the best indicator of academic achievement that could be obtained from the WISC-III. These researchers also concluded that using factor scores to help determine academic achievement becomes more complex and provides minimal results for both referred and nonreferred samples of students (Glutting et al., 1997).

*Summary*

The new WISC-IV has been recently updated and includes new core and supplemental subtests to its battery. These subtests have been deemed to provide information regarding clinical utility, but incremental validity research is lacking to determine if the new WISC-IV provides additional information regarding intervention strategies and methods. Many researchers highly recommend studies that involve incremental validity and it is especially important to include when examining the effectiveness of new subtests. Incremental validity can provide essential information about how new subtests contribute, improve, and/or enhance a test. The previous research discussed provides examples of how incremental validity has been examined in the WISC-III.

# CHAPTER III: SUMMARY AND DISCUSSION

The research has illustrated that incremental validity is an essential component to determine how new subtests or updated/revised tests improve or validate that they are measuring different domains of functioning or adding new and additional measures that were not previously assessed . However, research has also indicated that despite the need for incremental validity, many researchers neglect to measure this type of validity (Hunsley & Meyer, 2003). Although there is limited research regarding incremental validity, past studies have provided additional information that contributed to the incremental validity estimates for components of the WISC-III. The Watkins and Glutting (2000) study involving the WISC-III indicated that incorporating WISC-III subtest scatter and shape is not indicative of academic performance. Additionally, results from Glutting et al. (1997) study indicate that factor indexes do not provide additional information regarding the prediction of achievement that cannot already be provided by the full scale IQ in the WISC-III.

*Limitations of Literature Review*

This literature review has several limitations. There is a lack of research regarding how one could examine incremental validity in the WISC-III and there is no research available addressing incremental validity in the WISC-IV (Hunsley &Meyer, 2003). Consequently, the lack of research may have provided a biased view of examples of incremental validity. Lastly, because the literature review simply provides summarizations of past research in this area it does not contribute new information to the field of education.

*Limitations of Incremental Validity Studies*

There are several limitations of the research studies regarding incremental validity. In the first study by Watkins and Glutting (2000) their sample was limited and was not representative of the larger U.S. population. In the second study by Glutting et al. (1997) they only included a few assessment tools, and it would be useful to use additional cognitive and achievement tests to cross-validate ability and achievement measures. Additionally, it may be beneficial to determine if the factor scores correlate with other measures besides the factor scores. Furthermore, despite the effectiveness of the FSIQ to predict actual achievement, other factor scores may provide a better measurement of overall cognitive functioning for identification of difficulties or disabilities (Glutting et al., 1997).

*Implications for Future Research*

Additional research is strongly suggested to determine incremental validity in current and new assessment tools. It is important to determine how new subtests or completely new tests add and enhance measurements (Hunsley & Meyer, 2003). More research on the new WISC-IV will be essential to provide information on the new and updated core and supplemental tests. Additional studies could provide information about how these new subtests can contribute to interventions strategies and methods regarding special education decisions and services (Sattler & Dumont, 2004).

*Summary*

Incremental validity is an essential element to examine when determining the validity of new subtest or tests. Incremental validity can provide crucial information regarding the several aspects of one's performance or validity of test measurements.

Additionally, incremental validity could provide important information about what interventions or strategies may be useful depending on data from the specific assessment. Ultimately, this research has provided information about past literature on incremental validity and methods to examine it with the WISC-III. A critical analysis was provided that may help direct future research and practice implications.

# References

Esters, I. G., Ittenback, R. F., & Han, K. (1997). Today's IQ tests: Are they really better than their historical predecessors? *School Psychology Review,* 26, 2, 211-224.

Glutting, J.J., Youngstrom, E. A., Ward, T., Ward, S., & Hale, R. L. (1997). Incremental efficacy of WISC-III factor scores in predicting achievement: What do they tell us? *Psychological Assessment,* 9, 3, 295-301.

Horn & Noll. (1997). Human cognitive capabilities: Gf-gc theory. In D.P. Flanagan, J.L. Genshaft, & P. L. Harrison (Eds.) *Contemporary Intellectual Assessment: Theories, Tests, and Issues.* New York: Guilford Press.

Hunsley, J. & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment,* 15, 4, 446-455.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition, Technical Manual.* Itasca, IL: Riverside Publishing.

Sattler, J.M. (2001). Assessment *of children: Cognitive applications*. San Diego, CA: Sattler Publisher.

Sattler, J. M. & Dumont, R. (2004) *Assessment of children: WISC-IV and WPPSI-III supplement.* San Diego, CA: Sattler Publisher.

Watkins M. W. & Glutting, J. J. (2000). Incremental validity of WISC-III profile, elevation, scatter, and shape: Information for predicting reading and math achievement. *Psychological Assessment,* 12, 4, 402-408.

Wechsler, D. (2003). *Technical and interpretive manual.* San Antonio, TX: The Psychological Corporation.

Williams, P. D., Weiss, L. G., & Rolfus, E. (2003). *Theoretical model and test blueprint.* San Antonio, TX: The Psychological Corporation.