1-1-2009

Incremental Validity Of Wisc-Iv Factor Scores In Predicting Academic Achievement On The Wiat-Ii

Shalena R. Meis
Eastern Illinois University

THESIS MAINTENANCE AND REPRODUCTION CERTIFICATE

TO: Graduate Degree Candidates (who have written formal theses)

SUBJECT: Permission to Reproduce Theses

The University Library is receiving a number of request from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow these to be copied.

PLEASE SIGN ONE OF THE FOLLOWING STATEMENTS:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

Shalena R. Meis 8/4/09

Author's Signature Date

I respectfully request Booth Library of Eastern Illinois University **NOT** allow my thesis to be reproduced because:

____

____

____

Author's Signature Date

This form must be submitted in duplicate.

Incremental Validity of WISC-IV Factor Scores in Predicting

Academic Achievement on the WIAT-II

BY

Shalena R. Meis

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Specialist in School Psychology

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

2009
YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

8/3/09		8/3/09	
Dr. Gary Canivez THESIS COMMITTEE CHAIR	DATE	Dr. John Mace DEPARTMENT/SCHOOL CHAIR OR CHAIR'S DESIGNEE	DATE
8/3/09		8/3/09	
Dr. Michael Havey THESIS COMMITTEE MEMBER	DATE	Dr. Christine McCormick THESIS COMMITTEE MEMBER	DATE

Running Head: INCREMENTAL VALIDITY OF WISC-IV IN PREDICTING ACHIEVEMENT

Incremental Validity of WISC-IV Factor Scores in Predicting Academic Achievement on the WIAT-II

Shalena R. Meis

Eastern Illinois University

Abstract

This study examined the incremental validity of Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003) factor scores in predicting reading and math achievement on the Wechsler Individual Achievement Test-Second Edition (WIAT-II; Wechsler, 2002). Participants included K-12 students in a medium sized Midwestern school district who underwent an evaluation using the WISC-IV and the WIAT-II. Block entry multiple regression analyses were used to examine whether the WISC-IV factor scores substantially improved the prediction of reading or math achievement on the WIAT-II above and beyond the contribution made by the FSIQ. Results indicated that the WISC-IV FSIQ accounted for the majority of the variance in both reading and mathematics composite scores on the WIAT-II (34.3% and 55.7% respectively). In the prediction of reading achievement, the factor score indexes provided a statistically significant increment over the FSIQ, though the size of improvement (7.9%) was too small to be of clinical utility. In the prediction of math achievement, the factor scores did not provide a significant increment of prediction (1.1%) above and beyond the FSIQ. This study indicated that the WISC-IV factor scores did not substantially increase the predictive validity beyond the WISC-IV FSIQ. Therefore, like other investigations (Glutting et al., 1997; Glutting et al., 2006; Watkins et al., 2007), psychologists only need to interpret the WISC-IV FSIQ when predicting reading or math performance on the WIAT-II.

Acknowledgements

This study involved contributions from a variety of individuals, all who were pertinent to its completion. I would first like to thank my family and friends for their constant encouragement and support throughout the completion of this work. Of equal importance, I would like to acknowledge the individual under whom this study was conducted, Dr. Gary Canivez. Dr. Canivez provided guidance and advisement throughout the entire process. His interest in this study and support has been essential to creating a successful study.

In addition, I would like to acknowledge the members of my thesis committee, Dr. Michael Havey and Dr. Christine McCormick, for their constructive and encouraging guidance in directing this study.

Lastly, this study would not have been possible without the help of the members of our research team. I would like to thank Jonathon Cavanaugh, Emily Meyers, Jenny Kusy, Christi Blankenstein, Tiffany Miller, and Mark Valencia for all of the work that they contributed to this study.

I would again like to thank all the individuals involved in this study for their time and interest. Thank you for the guidance, support, and encouragement.

Table of Contents

Page

Abstract..3

Acknowledgements..4

List of Tables and Figures..6

Introduction...7

Method...28

Participants...28

Procedure...29

Instruments...30

Criteria...32

Data Analyses...32

Results...33

Discussion..35

References..39

List of Tables and Figures

Page

Table 1 Standard Score Descriptive Statistics for the WISC-IV and WIAT-II......32

Table 2 The Incremental Contributions of WISC-IV Factor Scores in Predicting Achievement Criteria on the WIAT-II..33

Incremental Validity of WISC-IV Factor Scores in Predicting Academic Achievement on the WIAT-II

Tests serve as indicators. They provide estimates of an individual's knowledge about a particular situation and may enable the prediction of an outcome. Mental ability tests serve the same purpose as all other tests. They sample behavior to predict a more extensive or important behavior or capability (Jensen, 1981).

In the early 1900s the concept of "intelligence test" became known. During the late 19th century, universal compulsory education was enacted in the United States and Europe. For the first time in history, large numbers of children whose parents did not have an education or who were not native English speakers began attending school. These students were exposed to curricula and academic standards that had been developed for a more privileged group of students, resulting in a high failure rate (Thorndike, 1997).

During this same time, Alfred Binet, a French psychologist became interested in the development of intelligence. In 1904, Binet became involved with the Free Society for the Psychological Study of the Child, a group of concerned parents and educational professionals. This group was particularly interested in the high failure rate experienced by many students enrolled in school. They identified two types of children who failed: those who could learn the material, but would not do so, and those who could not learn the material. At this time, the former group of children was labeled as "malicious," and the latter group of children was labeled as "stupid." The high failure rate was seen as a waste of resources, so the leaders of the period sought out ways to improve the effectiveness and efficiency of the schools by identifying the causes of school failure (Thorndike, 1997).

Alfred Binet and Theophile Simon, a psychiatrist, were commissioned by the French Ministry of Education to devise a practical and objective means for identifying mentally retarded children who could then be provided a specialized education according to their ability (Jensen, 1981). In 1905, Binet and Simon introduced a scale to measure what later became known as intelligence. The Binet-Simon scale consisted of 30 brief tasks arranged in order of difficulty. A child's intellectual level was defined by the most difficult task he or she could perform correctly (Thorndike, 1997). In 1908, a revised scale with items grouped by mental level or age was introduced, and, in 1911, the final version of the scale was published (Thorndike, 1990).

Although there were many other professionals involved in the measurement of intelligence, two individuals became particularly influential in measurement in the United States. Henry Goddard was the individual responsible for bringing Binet's test to the United States. He produced an American version of the test and promoted its use for various purposes. Lewis Terman, a professor at Stanford University, produced the Stanford revision of the Binet-Simon scale, which became known as the Stanford-Binet in its 1937 revision (Thorndike, 1997).

Since the development of the Simon-Binet scale, there have been many influential individuals involved in the measurement of intelligence. Though there have been numerous new developments in the field of intelligence since the early 1900s, the major application of intelligence tests today remains the same. Intelligence tests are used in schools today to determine whether a child's low level of academic achievement is due to a disability or some other cause (Thorndike, 1997).

Weiss and Prifitera (1995) stated that one of the most important functions of intelligence tests was their ability to predict student achievement. Intelligence tests have predictive validity, meaning that useful inferences can be drawn from the scores produced on the test. Thus, an individual's performance on the intelligence test can predict the same individual's performance in other situations or on a particular criterion, such as academic achievement (Jensen, 1981).

Correlations between intelligence and achievement are substantial (Naglieri & Bornstein, 2003). In fact, the median correlation between global intelligence and current achievement tests is .64. The correlation between intelligence and achievement can be even stronger when the two constructs are measured at the same time (Glutting, Youngstrom, Ward, Ward, & Hale, 1997). For example, Wechsler (1991) reported that correlations between the Full Scale IQ (FSIQ) on the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) and the achievement composites on the Wechsler Individual Achievement Test (WIAT; Wechsler, 1992) ranged from .70 to .81.

Although it is clear that the correlations between intelligence and achievement are significant, there has been considerable debate among professionals about whether IQ causes achievement, achievement causes IQ, or IQ and achievement are mutually influential. Brody (1997) and Ceci and Williams (1997) stated that the relationship between intelligence and achievement is reciprocal. IQ scores obtained prior to formal education are predictive of later academic achievements, yet IQ test scores change in response to the educational opportunities available to individuals.

Crano et al. (1972) reported "The use of intelligence tests is based on the assumption that such instruments tap a dimension distinct from the one measured in the achievement test-that intelligence is a prerequisite for achievement" (Crano et al., 1972, p. 260). Jensen (1998) stated that it is an unquestioned fact that IQ tests have a high degree of predictive validity for several educational criteria, including academic achievement.

Thorndike (1984) reported that 80 to 90 percent of the predictable variance in scholastic performance is accounted for by *g*, which is an abbreviation for the general intelligence factor, first identified by Spearman (1904). Jensen (1998) stated that highly *g*-loaded tests that do not resemble anything taught in school are only slightly less correlated with measures of scholastic performance than are the standard IQ and scholastic aptitude tests, which usually include some scholastic content. This indicates that the predictive validity of *g* does not depend on the test containing content that is taught in school. Jensen (1998) further stated that IQ tests predict achievement better than any other measurable variable because school learning is *g*-demanding. Students have to be able to learn new concepts, as well as transfer previously learned knowledge to the learning of new material.

On the other hand, Crano et al. (1972) reported that skills learned in specific settings generalize to more novel settings, and when a novel task is presented, the amount of available skills is greater if there is a large amount of past achievements. They continued, "…this year's specific learning achievements will generalize into next year's increased ability to solve novel problems, that is, into next year's intelligence. Intelligence would thus be viewed as a very general distillate of past achievements" (p. 262).

In an attempt to investigate this controversial issue, Crano et al. (1972) used the cross-lagged panel correlation technique to examine the relationship between IQ and achievement. The scores of 5,495 students who had taken intelligence and achievement tests in both fourth and sixth grades were analyzed to see whether a causal relationship existed between intelligence and achievement. In the overall sample, the predominant causal sequence was in the direction of intelligence causing later achievement. However, when the sample was divided into suburban and inner-city subunits, the intelligence causing achievement sequence only operated within the suburban sample. In fact, among the inner-city sample, the opposite held true. Despite this discrepancy, the researchers concluded that, "for the total group of respondents, the preponderant causal sequence is apparently in the direction of intelligence directly predicting later achievement to an extent significantly exceeding that to which achievement causes later intelligence" (p. 266).

Watkins, Lei, and Canivez (2006) noted that the Crano et al. (1972) study had several methodological problems, including violation of statistical assumptions, directional differences between subsamples, and the use of composite scales. The conclusions found by Crano et al. were weakened by reliance on group administered IQ and achievement tests. Crano et al. also relied on observed variables for their analyses. Observed variables are the factor-based ability scores that are obtained and interpreted during psychological testing (Glutting, Watkins, Konold, & McDermott, 2006). Observed variables contain measurement error. Therefore, relationships between observed variables can be biased by random errors of measurement. Lastly, Crano et al. used students from a single school district who were tested more than 40 years ago. Thus, the results found by Crano et al. should be interpreted with caution (Watkins et al., 2006).

Watkins et al. (2006) also examined the hypothesized causal influence of intelligence on achievement with a cross-lagged panel analysis of the WISC-III and achievement test scores of 289 students evaluated for special education eligibility with a test-retest interval of 2.8 years. Academic achievement was measured with a variety of tests, though the majority of the achievement test scores used in this study were from the Woodcock-Johnson Tests of Achievement (WJ-R; Woodcock & Johnson, 1989; WJ-III; Woodcock, McGrew, & Mather, 2001), WIAT, and Kaufman Test of Educational Achievement (K-TEA; Kaufman & Kaufman, 1985). Watkins et al. found that the paths from IQ scores at time 1 to IQ and achievement scores at time 2 were significant, whereas the paths from achievement scores at time 1 to IQ scores at time 2 were not significant. This finding suggested that intelligence influences or is related to future achievement, whereas achievement does not appear to substantially influence future intelligence.

It has been shown that intelligence predicts achievement, but does this remain true across racial/ethnic and gender groups? The literature provides evidence that well-constructed, properly standardized intelligence tests do, in fact, have comparative predictive validity among racial/ethnic and gender groups (Jensen, 1980).

Canivez reported that bias in psychological testing is the "extent to which there is systematic error in test scores as a function of a particular group membership" (Canivez, 2005, p. 68). Reynolds (1983) reported that test bias is evident when, "all ethnic or racial group differences on mental tests are due to inherent, artifactual biases embedded in the tests through flawed psychometric methodology" (p. 241). Though there are many issues related to test bias, the predictive validity bias of mental tests will be the issue discussed in the following paragraphs.

Predictive bias occurs when a particular test gives a biased prediction of the criterion for a particular minority group (Jensen, 1980). Neisser et al. (1996) reported that the relatively low mean of the distribution of African American intelligence test scores has been discussed for many years. Across studies, the African American mean is typically about one standard deviation below that of Caucasians (Jensen, 1980). Although there have been many reasons discussed for this difference, the relevant question here is whether intelligence tests have a predictive bias against African Americans.

Neisser et al. (1996) pointed out that such a bias would exist if the performance of African Americans on particular criterion variables (i.e. school achievement, college GPA, etc.) were systematically higher than the intelligence test scores would predict, but this is not the case. The actual regression lines for African Americans do not lie above those for Caucasians; there is even a slight tendency in the other direction (Jensen, 1980). Thus, considered as predictors of future performance, intelligence tests do not appear to be biased against African Americans (Neisser et al., 1996).

Reynolds and Nigl (1981) investigated the differential predictive validity of the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) for academic achievement assessed by the Wide Range Achievement Test (WRAT; Jastak & Jastak, 1976). The sample included 36 African American and 36 Caucasian children who had been referred to a private mental health clinic due to a variety of learning problems. Correlations were calculated between the WISC-R Verbal, Performance, and Full Scale IQs and the percentile scores on the WRAT subtests of Reading and Arithmetic. There were no significant differences found between the correlations or the regression equations, indicating a lack of bias among African American and Caucasian children. These researchers concluded that they found no evidence to support the cultural test bias hypothesis with regard to the predictive validity of intelligence tests for a population of inner city African American and inner city Caucasian children.

Naglieri and Hill (1986) compared the regression lines for the prediction of achievement test scores by the WISC-R FSIQ and the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983) Mental Processing Composite. The sample consisted of 86 African American and 86 Caucasian students. The results indicated that when using WISC-R or K-ABC IQs, the regression lines did not significantly differ in predicting group achievement test scores. The researchers also found no evidence of bias when WISC-R scores were used to predict K-ABC achievement scores. However, significant regression line differences were found when using the K-ABC Mental Processing Composite scores to predict K-ABC total achievement scores. The differences in the regression line intercepts suggested that using the Mental Processing Composite resulted in errors leading to the under prediction of African American achievement. Thus, the researchers supported the use of a common regression line in the prediction of achievement scores for African American and Caucasian children when the WISC-R FSIQ and the K-ABC Mental Processing Composite IQ are used to predict group achievement scores. They did not support the prediction of K-ABC achievement scores using the K-ABC Mental Processing Composite.

A study conducted by Weiss and Prifitera (1995) evaluated the relationship between WISC-III FSIQ and WIAT composite scores for reading, writing, mathematics, and language across racial (i.e. Caucasian, African American, and Hispanic) and gender groups. The results indicated that differential prediction was observed in 4 of the 12 comparisons made in this study. On the Reading Composite, for children with the same IQ score, Hispanics tended to score 2.0 points lower and females tended to score 1.8 points higher than would be expected from the common regression equation. On the Mathematics Composite, for children with the same IQ score, females tended to score 0.9 points higher than would be predicted by the common regression equation. On the Writing Composite, for children with the same IQ score, females tended to score 3.7 points higher than would be expected from the common regression equation. All effect sizes were too small to be of practical significance, and it was concluded that the WISC-III FSIQ adequately predicts achievement scores on the WIAT across racial/ethnic groups and genders.

Konold and Canivez (in press) replicated the study conducted by Weiss and Prifitera (1995) using the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003) and the Wechsler Individual Achievement Test-Second Edition (WIAT-II; Wechsler, 2002). The sample included participants from the standardization linking sample of the WISC-IV and WIAT-II ($N = 550$). The participants varied by demographic variables of gender, race/ethnicity (Caucasian, African American, and Hispanic), and parent education level (8-11 years, 12 years, 13-15 years, and 16 years). WISC-IV FSIQ and GAI scores were used to predict achievement scores on the WIAT-II. Results indicated no meaningful differences in WISC-IV predictions of WIAT-II scores across gender, race/ethnicity, or socioeconomic status. Although there may be slight differences in the predictive validity among racial and gender groups, subsequent research has supported Jensen's (1980) conclusion that most standard ability tests are not biased towards any group with respect to criterion validity.

Although a strong relationship between intelligence and achievement has been well documented in the literature, there have been few studies conducted evaluating the incremental validity of assessment measures (Hunsley & Meyer, 2003). Many definitions of incremental validity have appeared in the literature. Sechrest (1963) described incremental validity as increasing predictive efficiency; Wiggins (1973) described it as using psychological measures to improve on simple base-rate estimates; and Elliott, O'Donohue, and Nickerson (1993) described incremental validity as improving predictions in comparison with more or less expensive sources of data. Though there are several definitions of incremental validity, they all have in common the idea of relative predictive efficacy or an emphasis on the extent to which a measure adds to the prediction of a criterion above what can be predicted by other sources of data (Haynes & Lench, 2003).

Throughout the 1950s, the rate of theoretical and applied work on test validity rapidly increased (Hunsley & Meyer, 2003). Cronbach and Gleser (1957) reported that psychometricians began stating that new tests intended to be used for personnel decisions should be able to predict outcomes beyond the prediction of outcomes provided by the best available assessment procedures. However, it was Sechrest (1963) who first proposed the concept of incremental validity by stating that any psychological instrument intended for applied use must provide an improvement in prediction compared with other sources of data.

Meehl (1954) suggested that in order for a psychological test to have true utility it must lead to more effective treatment than would otherwise be provided. Sechrest (1963) stated that, at a minimum, a test should demonstrate incremental validity over brief case history information, biographical data, and brief interviews to have true utility in an applied context. Sechrest later suggested that psychological tests should improve prediction beyond what is possible with simpler and less costly tests to have true utility.

Although the concept of incremental validity was discussed in assessment texts, there was no systematic evaluation of incremental validity until Garb (1984) conducted a study examining the incremental validity of interview data, biographical data, personality tests, and neuropsychological tests (Hunsley & Meyer, 2003). Garb (1984) reported that biographical data, Minnesota Multiphasic Personality Inventory (MMPI) results, and neuropsychological test results all had incremental validity for assessing the psychological functioning of adults. However, interview data, sentence completion measures, the Rorschach, and the Thematic Apperception Test did not demonstrate incremental validity.

Garb (1984) reported that incremental validity data had not been translated into much actual research on incremental validity. He indicated that only 32 studies met his inclusion criteria and that many assessment measures and procedures had never been evaluated with an incremental validity paradigm. He concluded that little effort had been put forth to evaluate the incremental validity of psychological tests and assessment procedures.

In the past few decades, more psychologists have taken an interest in incremental validity research, resulting in greater knowledge about the concept. Hunsley and Meyer (2003) reported three conceptualizations of incremental validity research in the psychological literature. These conceptualizations included the incremental validity of test instruments, of test-informed clinical inferences, and of new measures. Since the present study focused on the incremental validity of test instruments, that will be the conceptualization reviewed in the following paragraphs.

The approach concerning the incremental validity of testing conducts nomothetic analyses in which information from a new source of data is evaluated in terms of its contribution to improving on the prediction of a clinically relevant criterion. This type of research focuses on adding new test data into a statistical equation in order to predict a criterion. Prediction is determined by the extent to which the sources of data can account for variance in a criterion. This approach provides evidence of either predictive validity or concurrent validity (Hunsley & Meyer, 2003).

A study by Kline, Snyder, Guilmette, and Castellanos (1992) provided a good example of this approach to the study of incremental validity. Within a sample of referred children, Kline et al. computed numerical indexes of profile elevation, shape, and variability for the WISC-R, the K-ABC, and the Stanford-Binet Intelligence Scale-Fourth Edition (SB-IV; Thorndike, Hagen, & Sattler, 1986). Earlier Cronbach and Gleser (1953) had found that subtest profiles provide three types of information: elevation, scatter, and shape. Elevation information is represented by a person's mean score over subtests.

Scatter information is how widely scores in that profile diverge from its mean. Shape information is where the ups and downs in the profile occur. Using hierarchical multiple regression analyses with achievement scores on the Wide Range Achievement Test-Revised (WRAT-R; Jastak, Jastak, & Wilkinson, 1984) as dependent variables, Kline et al. (1992) found that elevation was the most important predictor of achievement. Profile variability had no incremental validity, and profile shape had moderate unique predictive power for the WISC-R only.

Using a sample of 1,118 nondisabled and 538 disabled students, Watkins and Glutting (2000) used the elevation, scatter, and shape information from students' WISC-III subtest profiles to predict both reading and math achievement on the WJ-R. A series of regression analyses were conducted. The results were similar to those of Kline et al. (2002). Profile elevation was found to be a statistically significant predictor of achievement scores. Profile scatter did not increase the prediction of achievement scores. However, profile shape accounted for an additional 5%-8% of the variance in achievement scores.

Kaufman (1994) advocated for the "top down" hierarchical approach of test interpretation, meaning that factor scores are the next logical choice for interpretation after the FSIQ. Kaufman reported that factor scores capture dimensions of cognitive ability different from those reflected by the FSIQ or the individual subtests. Thus, ability constructs measured by factor deviation quotients might have a stronger association with certain achievement domains (Glutting et al., 1997).

The WISC-III factor scores include the Verbal Comprehension Index (VCI), the Perceptual Organization Index (POI), the Freedom from Distractibility Index (FDI), and the Processing Speed Index (PSI). Each factor index reflects different aspects of ability and is derived from separate subtest scores. Factor scores involve more than one correlated subtest, making them more reliable than individual subtests. During validation of the WISC-III, Wechsler (1991) found that the VCI showed higher average correlations with measures of overall achievement, as well as with more verbally oriented achievement criteria. The FDI demonstrated strong associations with individual measures of achievement as well. Conversely, the POI and the PSI showed the weakest associations with achievement.

Although factor score interpretation appeared to have benefits, the utility of factor scores had not been well researched, which is why Glutting et al. (1997) conducted a study to examine the incremental efficacy of WISC-III factor scores in predicting achievement on the WIAT. Using a sample of 283 nonreferred children and 636 children referred for evaluation, Glutting et al. used a series of hierarchical regressions to assess the prediction of achievement. The reading, math, writing, and language composites of the WIAT each served as the dependent measure in one set of regression analyses. The FSIQ was compared with the Verbal and Performance IQs, as well as the four factor scores (VCI, POI, FDI, and PSI) through block entry and removal within the hierarchical regressions.

Glutting et al. (1997) found that the factors provided a statistically significant increment, though the size of the increment was too small to be of clinical significance. The Verbal and Performance IQs and the four factor scores did not show any substantial increase in prediction of achievement after partialing out FSIQ. Thus, Glutting et al. concluded that the FSIQ is the best predictor of achievement on the WIAT.

Glutting, Watkins, Konold, and McDermott (2006) explored the incremental validity of factor scores once again by examining the utility of *observed* versus *latent* factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. Latent constructs underlie factor scores. They are perfectly reliable, though they cannot be observed directly. They are analyzed through structural equation modeling (SEM), which is a multivariate statistical technique used to identify relationships among latent variables. Observed scores are the factor-based ability scores that are obtained and interpreted during psychological testing. Observed scores contain error. However, most practitioners are more interested in observed scores because latent constructs are not observable and construct scores are very complex to compute. The four factor indexes in the WISC-IV are examples of observed factor scores. The WISC-IV factor indexes include the Verbal Comprehension Index (VCI), the Perceptual Reasoning Index (PRI), the Working Memory Index (WMI), and the Processing Speed Index (PSI).

Using the nationally stratified linking sample ($N = 498$) of the WISC-IV and WIAT-II, Glutting et al. (2006) used both hierarchical multiple regression analyses (MRA) and SEM to evaluate the importance of general versus specific abilities from the WISC-IV in predicting reading and math achievement. The MRA tested hypotheses about observed variables, whereas the SEM analyses tested hypotheses about latent variables.

At the observed variable level, the FSIQ accounted for approximately 60% of the variance in both reading and math composite scores. As was found in Glutting et al. (1997), the factor indexes provided a statistically significant increment over the FSIQ, but the size of the increment was too small to be of clinical utility. In fact, no single index provided an increment greater than 1% (Glutting et al., 2006).

At the latent variable level, only *g* and VC significantly influenced the reading and math achievement constructs. For both reading and math, *g* had a large effect size (.55 and .77 respectively). VC had an additional medium effect on reading (.37) and a small-to-medium effect on math (.17). Although both *g* and VC explain reading and math achievement, it is clear that *g* is the more powerful explanatory construct. Glutting et al. (2006) concluded that when using observed scores to predict reading and math achievement, the FSIQ is the only score that needs to be considered. When conducting explanatory research, both *g* and VC may be considered, though *g* usually accounts for the largest proportion of variance in achievement.

The MRA results of this study are lower than the 5%-16% increments found in the study conducted by Glutting et al. (1997) using the WISC-III and the WIAT linking sample. Thus, they concluded that the factor scores from the WISC-IV appear to have even less clinical utility in predicting achievement than the factor scores from the WISC-III. Although Detterman (2001) stated that *g* only predicts about 25%-50% of the variance in any particular outcome, leaving 50%-75% of the variance to be explained by other constructs, Glutting et al. (1997) agree more with Thorndike's (1985) statement that 80%-90% of predictable variance in criterion variables is accounted for by the single general score, or FSIQ, from an intelligence measure.

The literature clearly shows that the FSIQ is the best predictor of academic achievement, yet there are some who argue that the FSIQ becomes a less practical predictor of achievement when there is variability among the factor scores (Hale &

Fiorello, 2002). In fact, Hale and Fiorello (2002) stated that the FSIQ should never be interpreted when there is significant factor score variability.

Fiorello et al. (2007) conducted a study to examine the structure of intellectual functioning in children from three different clinical groups [i.e. Learning Disabled (LD), Attention-Deficit/Hyperactivity Disorder (ADHD), and Traumatic Brain Injury (TBI)]. Fiorello et al. used regression commonality analyses to estimate the size of the *g* factor. The assumption being that if the FSIQ is made up primarily of *g*, the highest-order commonality would be large relative to unique variance and lower-order commonalities. However, if the FSIQ has little common variance, the highest-order commonality would be small compared to lower-order commonalities and unique components.

Fiorello et al. (2007) reported that 50% of variance in FSIQ was due to commonalities among two or three of the indices, which they referred to as interpretable shared variance components. The main conclusion presented by Fiorello et al. followed along the same lines as that presented by Hale and Fiorello (2002). Fiorello et al. (2007) stated that they found evidence to support idiographic Index interpretation over nomothetic interpretation of the FSIQ for certain populations (i.e. LD, ADHD, and TBI). Based on their study, they concluded that the global interpretation of the FSIQ is inappropriate because the shared variance among the four indices is small and it may obscure important aspects of cognitive functioning for children in clinical groups. They further stated that interpretation at the Index level is appropriate because the scores are reliable and stable and provide substantial portions of FSIQ variance.

Faust (2007) commented on the statement presented by Fiorello et al. (2007) suggesting that because the FSIQ cannot account for all of the variance in an outcome other constructs should be taken into account. Faust stated that given the randomness of human behavior and measurement error, the failure to account for all of the variance does not mean that there is a considerable amount of remaining variance that needs to be accounted for. He further stated that, in predictive domains, adding valid variables to other valid variables may not increase, and may even decrease, accuracy. He concluded by stating that validity is often not cumulative, even though one or a few variables may provide far from complete predictive power. He provided the example that if Variable A is a powerful predictor and Variable B is a weaker predictor and the two lead to different conclusions, it is generally best to go with Variable A over Variable B and not attempt to integrate the two.

Dana and Dawes (2007) used simulation to construct a null result against which to compare the results presented by Fiorello et al. (2007). Dana and Dawes created 10 "subtest" scores that were defined on a single factor with error and then assigned to four arbitrary indices that summed to a FSIQ. They removed all scores that were ±1⅓ standard deviations from the mean and performed commonality or shared variance analyses. This process was repeated several times to determine sampling variability. The simulated FSIQs were composed of largely unique contributions (over 45% of variance) by the four indices. The commonality of all four indices, representing *g*, contributed just 4% of the variance on average. Furthermore, in 40% of the trials, the commonality of the indices explained less than 1.7% of the variance, the smallest commonality that Fiorello et al. found in their study.

Dana and Dawes (2007) demonstrated that a dataset can produce relatively small higher-order commonalities and relatively large uniqueness and lower-order commonalities. Thus, the assumption made by Fiorello et al. (2007) that lower-order commonalities represent sources of variance distinct from *g* is false because *g* was the only possible cause of correlations in that particular dataset. Dana and Dawes (2007) concluded that the results produced by Fiorello et al. (2007) do not provide evidence to support their claims that FSIQ interpretation is inappropriate or that lower order variance components are interpretable.

Watkins, Glutting, and Lei (2007) conducted a study to determine whether the argument that FSIQ scores are invalid predictors of achievement in the presence of significant factor score variability is actually valid. The sample included 412 children from the WISC-III/WIAT linking sample, 136 children from the WISC-IV/WIAT-II linking sample, and 460 children enrolled in special education programs. The participants with and without statistically significant factor score variability were matched on FSIQ, age, gender, race/ethnicity, and disability category. Moderated multiple regression analyses were used to detect any bias in the predictive validity of FSIQ scores between participants with and without significant factor score variability.

The results found by Watkins et al. (2007) indicated that the FSIQ is a significant predictor of reading and math achievement, regardless of factor score variability. Neither factor score variability, nor the interaction of the FSIQ and factor score variability, made a statistically significant incremental contribution to the prediction of reading and math achievement scores. The hypothesis that the FSIQ is invalid in the presence of factor score variability was refuted, and, thus, the FSIQ remained the best predictor of academic achievement, regardless of variability in the factor scores.

Schneider (2008) also commented on the Fiorello et al. (2007) study. He stated that the arguments presented by Fiorello et al. (2007) were based on false assumptions and logical errors. Schneider (2008) used simulated data to show that Fiorello et al.'s (2007) use of commonality analysis can lead to incorrect conclusions. Schneider (2008) stated that commonality analysis is a form of regression analysis in which all possible orders of entry of the predictor variables are considered. It identifies the amount of utility each predictor variable has in predicting an outcome variable, which enables researchers to know which predictors can safely be dropped.

Schneider (2008) stated that Fiorello et al. (2007) used commonality analysis for purposes for which it is not suited. He further stated that Fiorello et al. (2007) failed to check their assumptions about commonality analysis, leading to a misinterpretation of results. Simulated datasets (Dana & Dawes, 2007) have shown that when Fiorello et al.'s (2007) methods are followed, the highest-order commonality is a poor index of the size of *g* and the lower-order commonalities and uniqueness are contaminated by *g*. Schneider (2008) stated that the reason for this is that when predictors are imperfect indicators of *g*, the uniqueness in each variable will contain some *g* variance that the other variables failed to capture.

Schneider (2008) also commented that although Fiorello et al. (2007) cite Pedhazur (1997) as their source of computational methods for commonality analysis, they failed to recognize Pedhazur's warnings that commonality analysis is useful for predictive but not explanatory purposes. Schneider (2008) concluded by stating that, "because the same results can come from almost any model of intelligence, drawing strong conclusions based on a commonality analysis has all the dangers of being led by the same forces that drive Ouija boards" (p. 7).

Thus, although Hale and Fiorello (2002) and Fiorello et al. (2007) attempted to discredit FSIQ interpretation, there are many expert researchers in the field that have empirically demonstrated the global interpretation of the FSIQ not only appropriate, but necessary.

Meyer et al. (2001) stated that the incremental validity of an assessment measure must be evaluated to ensure that the use of that instrument is empirically sound and justifiable. Incremental validity supplements traditional psychometric dimensions of content, convergent, predictive, discriminant, and other forms of validity. Incremental validity provides information that is useful to professionals when evaluating the contribution of a new measure or when selecting the most appropriate measure for the case (Haynes & Lench, 2003).

Hunsley and Meyer (2003) reported that a major challenge for psychological testing and assessment research on incremental validity is the noncumulative nature of the published research. They argued that there has been too little attention paid to how clinically meaningful an increment in validity is, and further stated that greater attention should be focused on the need for replication of incremental validity research. Hunsley and Meyer stressed the importance of conducting incremental validity studies by proposing that future studies should attempt to conceptually replicate previous findings by using similar order of entry strategies for variables in multiple regression analyses.

It is due to this need that the present study was conducted. This study served as a replication of previous studies of similar nature (Kline et al., 1992; Glutting et al., 1997;

Konold, 1999; Watkins & Glutting, 2000; Glutting et al., 2006; Watkins et al., 2007). Limited studies on incremental validity have included the WISC-IV. The WISC-IV is the latest edition of the Wechsler intelligence test series and commonly used in school psychology practice. The present study sought to replicate previous findings using the WISC-IV with a different sample. It examined the incremental validity of WISC-IV factor scores in predicting achievement on the reading and math composites of the WIAT-II.

Method

Participants

The participants in the present study included K-12 students in a medium sized Midwestern school district. All students were referred for a psychological evaluation and underwent assessment as the result of a suspected disability. The sample included students who were evaluated between 2003 and 2008. As of 2008, the total enrollment of the school district was 9,326. The school population was comprised of 4,252 Caucasian students (45.6%), 3,543 African American students (38.0%), 621 Hispanic students (6.7%), 881 Asian/Pacific Islander students (9.4%), and 29 students classified as Other (0.3%). The district's low income rate was 43.9% and the limited English proficiency rate was 4.3%.

Data in the present study were collected via records review from the special education files of those students referred for a psychological evaluation for the determination of the presence of a disability. Demographic information including age, gender, race/ethnicity, family status, and language spoken were obtained from each student's special education folder. Permission to access this information was granted by the district's School Board and the Special Education Coordinator.

The final sample included data on 145 students who were evaluated with the WISC-IV and the WIAT-II. Some of the demographic information on certain students was unknown. Data on age, grade, and sex was reported for 144 students, and data on race and disability was reported for 143 students. The age range for the sample was six years old to fifteen years old (*M* = 9.98, *SD* = 1.95). The grade levels included in the sample ranged from first grade to eighth grade. There were 93 male (63.7%) and 51 female (34.9%) students. Of the students that were included in the sample, 83 were African American (56.8%), 51 were Caucasian (34.9%), 8 were biracial/multiracial (5.5%), and 1 was Hispanic (.7%). It is important to note that the demographic makeup of the sample was not consistent with the demographics of the school district. The sample included a higher percentage of African American and biracial/multiracial students than would be expected according to the demographics of the school district. The sample consisted of 107 students with a Specific Learning Disability (73.3%), 15 students not disabled (10.3%), 11 students with Mental Retardation (7.5%), 5 students with a Severe Emotional Disability (3.4%), 3 students with Autism (2.1%), 1 student with a Speech or Language Impairment (.7%), and 1 student with Traumatic Brain Injury (.7%).

Procedure

From 2007 to 2008, a research team from Eastern Illinois University collected the data utilized in the present study. Information was recorded from each student's special education file onto a worksheet. The information used in the study included student identification number, number of evaluations, date of birth, date of test, chronological age, grade, sex, race/ethnicity, disability, cognitive test used during the evaluation, cognitive test composites, cognitive subtest scores, achievement tests used during the evaluation, achievement composite scores, and achievement subtest scores.

This information was entered into a computer database by trained undergraduate psychology majors. Student names and personal information were anonymous. The student identification number was used to reference students.

Instruments

The Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003) measures cognitive abilities among children 6 years 0 months through 16 years 11 months. The battery includes 15 subtests (*M*s = 10, *SD*s = 3), 10 of which form the core battery. The WISC-IV FSIQ is comprised of the 10 subtests that make up the core battery. The 10 core subtests are organized to form four factor-based indexes: the Verbal Comprehension Index (VCI), the Perceptual Reasoning Index (PRI), the Working Memory Index (WMI), and the Processing Speed Index (PSI). The VCI is made up of Similarities, Vocabulary, and Comprehension subtests. The PRI includes Block Design, Picture Concepts, and Matrix Reasoning subtests. The WMI is comprised of Digit Span and Coding subtests, and the PSI is made up of Letter-Number Sequencing and Symbol Search subtests. The five subtests excluded from the core battery are referred to as supplemental subtests, and each is associated with one of the four indexes: VCI (Information and Word Reasoning subtests), PRI (Picture Completion subtest), WMI (Arithmetic subtest), and PSI (Cancellation subtest). Each of the four indexes is expressed as a standard score (*M* = 100, *SD* = 15) (Wechsler, 2003).

The WISC-IV was standardized on a sample of 2,200 children who were chosen to closely match the 2002 U.S. census data on the variables of age, gender, geographic region, race/ethnicity, and socioeconomic status. The VCI subtests generally have the highest *g* loadings, followed by the PRI, WMI, and PSI subtests. The average internal consistency coefficients are .94 for VCI, .92 for PRI, .92 for WMI, .88 for PSI, and .97 for FSIQ. The FSIQ and factor indexes reliabilities are usually high (>.90+), whereas the subtest reliabilities are generally medium (.80-.89). The WISC-IV is a stable instrument over the short term with average test-retest coefficients of .93 for VCI, .89 for PRI, .86 for WMI, .86 for PSI, and .93 for FSIQ (Wechsler, 2003).

The Wechsler Individual Achievement Test-Second Edition (WIAT-II; Wechsler 2002) measures academic achievement among individuals aged 4 to 85 years old. It contains nine subtests that combine to form four composite scores: Reading, Mathematics, Written Language, and Oral Language (*M*s = 100, *SD*s = 15). The Reading Composite is composed of Word Reading, Reading Comprehension, and Pseudoword Decoding subtests. The Mathematics Composite is made up of Numerical Operations and Math Reasoning subtests. The Written Language Composite is formed from Spelling and Written Expression subtests. The Oral Language Composite is comprised of Listening Comprehension and Oral Expression subtests (Wechsler, 2002).

The WIAT-II was standardized on 5,586 individuals aged 4-0 to 19-11 and in grades pre-kindergarten to 12. The standardization sample closely approximated the 1998 U.S. census on grade, age, gender, race/ethnicity, geographic region, and socioeconomic status. Data on internal consistency, test-retest stability, and interscorer reliability demonstrated a high level of precision. The correlations for the composite scores ranged from .87 to .99 (Wechsler, 2002).

Criteria

All of the participants in the present study completed the WISC-IV and the WIAT-II. The present study focused on outcomes in reading and math. Therefore, the participants completed the three subtests of the Reading Composite (i.e. Word Reading, Reading Comprehension, and Pseudoword Decoding) and the two subtests of the Mathematics Composite (i.e. Numerical Operations and Math Reasoning) on the WIAT-II.

Data Analyses

The contributions of the different factor scores on the WISC-IV to prediction of achievement criteria on the WIAT-II were evaluated through a series of hierarchical regression analyses. The WIAT-II Reading Composite and Mathematics Composite each served as criteria in separate regression analyses. The contribution of the WISC-IV FSIQ was compared with the four factor scores (VCI, PRI, WMI, PSI) through block entry and removal within the hierarchical multiple regression analyses.

The FSIQ was entered into the regression model by itself in the first block, and the four factor scores were entered as a group in the second block. The change in explained achievement variance from the entrance of the second block with the factor scores provided an estimate of the maximum predictive increment possible through the use of the factor scores in addition to the FSIQ.

Results

Block entry multiple regression analyses were used to address the research question: Do the WISC-IV factor index scores significantly and substantially improve the prediction of WIAT-II reading or math achievement above and beyond the contribution made by the WISC-IV FSIQ? Table 1 presents the descriptive statistics for the WISC-IV and WIAT-II. Table 2 presents the percentage of variance of WIAT-II Reading and WIAT-II Mathematics that was accounted for by WISC-IV FSIQ and the addition of WISC-IV factor index scores. Table 2 presents the improvement obtained by entering the four factor scores into the model after first entering FSIQ. The change in explained achievement variance provides an estimate of the maximum predictive increment possible through the use of factor scores in addition to the FSIQ. Table 2 also presents the unique contributions of each of the factor index scores (VCI, PRI, WMI, and PSI) when all variables were simultaneously included in the regression equation.

Table 1

Standard Score Descriptive Statistics for the WISC-IV and WIAT-II (N = 145)

	Range	*M*	*SD*
WISC-IV			
FSIQ	48-123	85.10	13.47
VCI	59-116	87.96	12.49
PRI	51-143	90.38	15.24
WMI	50-120	85.26	13.05
PSI	50-123	88.47	14.58
WIAT-II			
Reading Composite	41-119	80.54	11.70
Mathematics Composite	40-135	83.06	15.97

Note. WISC-IV = *Wechsler Intelligence Scale for Children-Fourth Edition* (Wechsler, 2003); WIAT-II = *Wechsler Individual Achievement Test-Second Edition* (Wechsler, 2002); FSIQ = Full Scale IQ; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index

Reading

The first analysis of WISC-IV factors improvement of prediction of the WIAT-II Reading Composite was statistically significant, R^2 change = .08, $F(4, 139) = 4.73$, $p = .001$. Although statistically significant, the increment of change was small. The WISC-IV FSIQ accounted for 34.3% of the variance in WIAT-II Reading. As a group, the four factors explained an additional 7.9% of the variance in WIAT-II Reading. The unique contributions of each factor score were also small, ranging from 0.1% to 1.2% of the WIAT-II Reading variance. The VCI contributed the least amount of variance (0.1%), the PRI contributed 0.3% of variance, the WMI contributed 0.7% of variance, and the PSI contributed the most amount of variance (1.2%).

Mathematics

The second analysis of WISC-IV factors improvement of prediction of the WIAT-II Mathematics Composite was not statistically significant, R^2 change = .01, $F(4, 139) = .92$, $p = .46$. The WISC-IV FSIQ accounted for 55.9% of the variance in WIAT-II Math. The WISC-IV factor indexes only explained an additional 1.1% of the variance in WIAT-II Mathematics. The unique contributions of each factor score were small, ranging from 0.6% to 1.0% of the WIAT-II Mathematics variance. These results were similar to that of reading achievement. The VCI contributed the least amount of variance (0.6%), the PRI contributed 0.7% of variance, the WMI contributed 0.9% of variance, and the PSI contributed the greatest amount of variance (1.0%).

Table 2

The Incremental Contributions of WISC-IV Factor Scores in Predicting Achievement Criteria on the WIAT-II

	WIAT-II Reading		WIAT-II Mathematics	
Predictor	% Variance	% Increment[a]	% Variance	% Increment[a]
FSIQ	34.3*	34.3*	55.9*	55.9*
Four Factors (*df* = 4)[b]	42.2	7.9*	57.0	1.1
VCI		0.1		0.6
PRI		0.3		0.7
WMI		0.7		0.9
PSI		1.2		1.0

Note. WISC-IV = *Wechsler Intelligence Scale for Children-Fourth Edition* (Wechsler, 2003); WIAT-II = *Wechsler Individual Achievement Test-Second Edition* (Wechsler, 2002); FSIQ = Full Scale IQ; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index. [a]Unless indicated otherwise, all unique contributions are squared part correlations, equivalent to the change in R^2 if this variable were entered last in a block entry regression. [b]Partialing out FSIQ.
* $p < .05$

Discussion

The present study served as a replication of previous studies examining the incremental validity of cognitive assessments in predicting achievement (Kline et al., 1992; Glutting et al., 1997; Konold, 1999; Watkins & Glutting, 2000; Glutting et al., 2006; Watkins et al., 2007). This study specifically examined the incremental efficacy of the WISC-IV factor scores in predicting reading and math achievement on the WIAT-II and whether the WISC-IV factor index scores should be used in interpretation.

The findings from the present study indicated that the WISC-IV FSIQ accounted for the majority of the variance in both reading and mathematics composite scores on the WIAT-II (34.3% and 55.7% respectively). These results were extremely similar to the results found by Glutting et al. (1997) involving the WISC-III and WIAT. The results of the referred sample indicated that the WISC-III FSIQ accounted for 36.0% of the WIAT Reading variance and 55.5% of the WIAT Mathematics variance.

In the prediction of WIAT-II reading achievement, the WISC-IV factor index scores provided a statistically significant increment over the WISC-IV FSIQ, though the size of improvement (7.9%) was too small to be of clinical utility. Glutting et al. (1997) found a similar result in the prediction of WIAT reading achievement. In the referred sample, the WISC-III factor index scores provided 8.8% increment over the WISC-III FSIQ.

In the prediction of WIAT-II math achievement, the results from the present study indicated that the WISC-IV factor scores did not provide a significant increment of prediction (1.1%) above and beyond the WISC-IV FSIQ. This result was similar to the result found by Glutting et al. (2006) indicating that the WISC-IV factor scores provided 0.3% increment of prediction over the WISC-IV FSIQ in the prediction of WIAT-II math achievement.

The proportion of variance predicted for reading achievement (34.3%) was less than the proportion of variance predicted for math achievement (55.7%), perhaps due to the makeup of the sample. The sample used in the present study included 107 participants diagnosed with a Specific Learning Disability (73.3%). Of the participants diagnosed with a Specific Learning Disability, 32 students were diagnosed with a Specific Learning Disability in reading (21.9%), 18 students were diagnosed with a Specific Learning Disability in math (12.3%), and 48 students were diagnosed with a Specific Learning Disability in both reading and math (32.9%). This may have had an effect on the predictive relationship between IQ and achievement. By definition, students with a Specific Learning Disability in reading achieve lower in reading than expected or predicted by intellectual ability. Since there were somewhat more students with a

Specific Learning Disability in reading than in math, the correlation between IQ and reading achievement may have been lower than the correlation between IQ and math achievement as a result. Thus, the proportion of achievement variance accounted for by the FSIQ was less for reading than for math.

The individual WISC-IV factor scores did not provide a significant increase in prediction of WIAT-II achievement after partialing out WISC-IV FSIQ. In the prediction of WIAT-II reading and math achievement, no single index provided an increment greater than 1.2%. This result was similar to the findings by Glutting et al. (1997) and Glutting et al. (2006). In the referred sample, the unique contribution of each WISC-III factor score ranged from 0-4.4% in the prediction of WIAT reading and math achievement (Glutting et al., 1997). Glutting et al. (2006) found that none of the WISC-IV factor scores provided an increment greater than 2% in the prediction of WIAT-II reading and math achievement.

The Verbal Comprehension Index provided the weakest contribution in the prediction of reading and math achievement (0.1% and 0.6% respectively). Glutting et al. (1997) suggested that these results may reflect the fact that the VCI is so highly correlated with the FSIQ that it is difficult for the VCI to make a unique contribution as an independent variable. The Processing Speed Index contributed the largest increment of prediction (1.2% for reading achievement and 1.0% for math achievement). Although the PSI contributed the largest increment, the variance was too small to provide clinically useful information. The Perceptual Reasoning Index and the Working Memory Index did not demonstrate any substantial increment with the achievement composites.

This study had several limitations. First, these data were archival data. Several school psychologists employed by the school district where these data were obtained were the individuals who actually evaluated the children and documented the data. While data were strictly copied from student files, the accuracy of the data in the files cannot be assured. Second, the sample size was fairly small. A majority of data available were of the WISC-III and the WIAT. Since the WISC-IV and the WIAT-II are the newest versions of these assessments, the number of evaluations using both of these instruments was limited, which produced a smaller sample size. Third, these data were collected from only one school district and geographic location. As previously stated, the sample used in this study was not representative of the school district or current census data. Thus, results cannot be generalized to dissimilar students.

Despite these limitations, the findings from the present study were similar to the results found in previous studies involving the linking sample of the WISC-III and WIAT (Glutting et al., 1997) and the linking sample of the WISC-IV and WIAT-II (Glutting et al., 2006). This study provided evidence suggesting that factor score interpretation did not appear to be clinically relevant in predicting reading or math achievement on the WIAT-II. The WISC-IV factor scores did not substantially increase the predictive validity beyond WISC-IV FSIQ. Therefore, it is recommended that school psychologists need only interpret the WISC-IV FSIQ when predicting reading or math performance on the WIAT-II.

References

Brody, N. (1997). Intelligence, schooling, and society. *American Psychologist*, 52, 1046-1050.

Canivez, G. L. (2005). Bias (testing). In S. W. Lee (Ed.), *The encyclopedia of school psychology* (pp. 67-69). Thousand Oaks, CA: Sage Publications.

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist*, 52, 1051-1058.

Crano, W. D., Kenny, D. A., & Campbell, D. T. (1972). Does intelligence cause achievement?: A cross-lagged panel analysis. *Journal of Educational Psychology*, 63, 258-275.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin*, 50, 456-473.

Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions*. Urbana: University of Illinois Press.

Dana, J., & Dawes, R. (2007). Comment on Fiorello et al. "Interpreting intelligence test results for children with disabilities: Is global intelligence relevant?" *Applied Neuropsychology*, 14, 21-25.

Detterman, D. K. (2001). General intelligence: Cognitive and biological explanations. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp.223-243). Mahwah, NJ: Lawrence Erlbaum Associates.

Elliott, A. N., O'Donohue, W. T. O., & Nickerson, M. A. (1993). The use of sexually anatomically detailed dolls in the assessment of sexual abuse. *Clinical Psychology Review*, 13, 207-221.

Faust, D. (2007). Some global and specific thoughts about some global and specific issues. *Applied Neuropsychology*, 14, 26-36.

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology*, 14, 2-12.

Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review*, 4, 641-655.

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education*, 40, 103-114.

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, 295-301.

Hale, J. B., & Fiorello, C. A. (2002). Beyond the academic rhetoric of *g*: Intelligence testing guidelines for practitioners. *NASP Communiqué*, 31(2), 1-8.

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment*, 15, 456-466.

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

Jastak, J. F., & Jastak, S. (1976). *Wide Range Achievement Test*. Wilmington, DE: Jastak Associates.

Jastak, J. F., Jastak, S., & Wilkinson, G. S. (1984). *Wide Range Achievement Test-Revised*. Wilmington, DE: Jastak Associates.

Jensen, A. R. (1980). *Bias in mental testing*. New York: The Free Press.

Jensen, A. R. (1981). *Straight talk about mental tests*. New York: The Free Press.

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

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: John Wiley & Sons.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Services.

Kaufman, A. S., & Kaufman, N. L. (1985). *Kaufman Test of Educational Achievement*. Circle Pines, MN: American Guidance Services.

Kline, R. B., Snyder, J., Guilmette, S., & Castellanos, M. (1992). Relative usefulness of elevation, variability, and shape information from WISC-R, K-ABC, and Fourth Edition Stanford-Binet profiles in predicting achievement. *Psychological Assessment*, 4, 426-432.

Konold, T. R. (1999). Evaluating discrepancy analyses with the WISC-III and WIAT. *Journal of Psychoeducational Assessment*, 17, 24-35.

Konold, T. R. & Canivez, G. L. (in press). Differential relationships among WISC-IV and WIAT-II scales: An evaluation of potentially moderating child demographics. *Educational and Psychological Measurement*.

Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R. et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128-165.

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psychoeducational Assessment*, 21, 244-260.

Naglieri, J. A., & Hill, D. S. (1986). Comparison of WISC-R and K-ABC regression lines for academic prediction with black and white children. *Journal of Clinical Child Psychology*, 15, 352-355.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J. et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77-101.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed., pp. 199-211). Orlando, FL: Holt, Rinehart, & Winston.

Reynolds, C. R. (1983). Test bias: In God we trust; All others must have data. *The Journal of Special Education*, 17, 241-260.

Reynolds, C. R., & Nigl, A. J. (1981). A regression analysis of differential validity in intellectual assessment for black and for white inner city children. *Journal of Clinical Child Psychology*, 10, 176-179.

Schneider, W. J. (2008). Playing statistical ouija board with commonality analysis: Good questions, wrong assumptions. *Applied Neuropsychology*, 15, 44-53.

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement*, 23, 153-158.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

Thorndike, R. L. (1984). *Intelligence as information processing: The mind and the computer*. Bloomington, IN: Center on Evaluation, Development, and Research.

Thorndike, R. L. (1985). The central role of general ability in prediction. *Multivariate Behavioral Research*, 20, 241-254.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition*. Chicago: Riverside.

Thorndike, R. M. (1990). Origins of intelligence and its measurement. *Journal of Psychoeducational Assessment*, 8, 223-230.

Thorndike, R. M. (1997). The early history of intelligence testing. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*. New York: Guilford.

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, 402-408.

Watkins, M. W., Glutting, J. J., & Lei, P. (2007). Validity of the full-scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, 14, 13-20.

Watkins, M. W., Lei, P., & Canivez, G. L. (2006). Psychometric intelligence and achievement: A cross-lagged panel analysis. *ScienceDirect*, 35, 59-68.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1992). *Wechsler Individual Achievement Test*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2002). *Manual for the Wechsler Individual Achievement Test-Second Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2002). *Wechsler Individual Achievement Test-Second Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Manual for the Wechsler Intelligence Scale for Children-Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Weiss, L. G., & Prifitera, A. (1995). An evaluation of differential prediction of WIAT achievement scores from WISC-III FSIQ across ethnic and gender groups. *Journal of School Psychology*, 33, 297-304.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Menlo Park, CA: Addison-Wesley.

Woodcock, R. W. & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery-Revised*. Allen, TX: DLM.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.