WISC-IV Matrix Reasoning and Picture Concepts Subtests: Does the Use of Verbal Mediation Confound Measurement of Fluid Reasoning?

Jaime L. Houskeeper

Pacific University
WISC-IV Matrix Reasoning and Picture Concepts Subtests: Does the Use of Verbal Mediation Confound Measurement of Fluid Reasoning?

Abstract
The WISC-IV introduced two new subtests, Matrix Reasoning (MR) and Picture Concepts (PC). The present study intends to address the need for independent evaluation of construct validity, particularly as it relates to reports that both of these presumed nonverbal reasoning measures may be impacted by verbal mediation and verbal skills. The present study included a community sample of 76 children ranging in age from 6 to 16 years. It was primarily hypothesized that MR and PC would be more closely related to a measure of verbal abilities from the WISC-IV (Similarities subtest) than to a measure of visuo-spatial skills, Beery's VMI. These hypotheses were not supported, however, and a discussion of strengths and weaknesses of this study, as well as implications for future research, is presented. Specifically, the present results suggest that the Matrix Reasoning subtest in particular may be impacted by visual abilities more than verbal abilities.

Degree Type
Dissertation

Degree Name
Doctor of Psychology (PsyD)

Committee Chair

Subject Categories
Psychiatry and Psychology

Comments
Library Use: LIH

This dissertation is available at CommonKnowledge: https://commons.pacificu.edu/spp/212
Copyright and terms of use

If you have downloaded this document directly from the web or from CommonKnowledge, see the “Rights” section on the previous page for the terms of use.

If you have received this document through an interlibrary loan/document delivery service, the following terms of use apply:

Copyright in this work is held by the author(s). You may download or print any portion of this document for personal use only, or for any use that is allowed by fair use (Title 17, §107 U.S.C.). Except for personal or fair use, you or your borrowing library may not reproduce, remix, republish, post, transmit, or distribute this document, or any portion thereof, without the permission of the copyright owner. [Note: If this document is licensed under a Creative Commons license (see “Rights” on the previous page) which allows broader usage rights, your use is governed by the terms of that license.]

Inquiries regarding further use of these materials should be addressed to: CommonKnowledge Rights, Pacific University Library, 2043 College Way, Forest Grove, OR 97116, (503) 352-7209. Email inquiries may be directed to: copyright@pacificu.edu
WISC-IV MATRIX REASONING AND PICTURE CONCEPTS SUBTESTS: DOES THE USE
OF VERBAL MEDIATION CONFOUND MEASUREMENT OF FLUID REASONING?

A DISSERTATION

SUBMITTED TO THE FACULTY

OF

SCHOOL OF PROFESSIONAL PSYCHOLOGY

PACIFIC UNIVERSITY

HILLSBORO, OREGON

BY

JAIME L. HOUSKEEPER, M.S.

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF CLINICAL PSYCHOLOGY
Abstract

The WISC-IV introduced two new subtests, Matrix Reasoning (MR) and Picture Concepts (PC). The present study intends to address the need for independent evaluation of construct validity, particularly as it relates to reports that both of these presumed nonverbal reasoning measures may be impacted by verbal mediation and verbal skills. The present study included a community sample of 76 children ranging in age from 6 to 16 years. It was primarily hypothesized that MR and PC would be more closely related to a measure of verbal abilities from the WISC-IV (Similarities subtest) than to a measure of visuo-spatial skills, Beery’s VMI. These hypotheses were not supported, however, and a discussion of strengths and weaknesses of this study, as well as implications for future research, is presented. Specifically, the present results suggest that the Matrix Reasoning subtest in particular may be impacted by visual abilities more than verbal abilities.
# Table of Contents

Introduction.........................................................................................................................1  

Literature Review..................................................................................................................4  

Methods...............................................................................................................................26  

Results.................................................................................................................................30  

Discussion............................................................................................................................37  

References...........................................................................................................................45  

Appendices..........................................................................................................................49
List of Tables

Table 1: CHC Abilities and Related WISC-IV Subtests

Table 2: Structure of WISC-IV

Table 3: Subtest Score Distributions

Table 4: Degree of Skewness and Kurtosis for Key Variables

Table 5: Simple Linear Regression Model with Perceptual Reasoning Subtests as a Function of Similarities Scores

Table 6: Correlation Coefficients for Subtests

Table 7: T-values for Correlation Comparisons
List of Figures

Figure 1: Example item from Matrix Reasoning subtest of the WISC-IV……………………...56

Figure 2: Example item from Picture Concepts subtest of the WISC-IV……………………….57
WISC-IV Matrix Reasoning and Picture Concepts subtests:

Does the use of verbal mediation confound measurement of fluid reasoning?

Within the field of psychology, intellectual assessment remains a relatively young discipline. As such, its tools are in a state of constant growth and development. Recently, with the advent of research- and theory-informed assessment (Flanagan & Kaufman, 2004), major revisions to intelligence tests continue to occur at an increasing pace. The Wechsler scales are revised at shorter intervals with each version. When evaluating existing intellectual instruments, it is important to consider intellectual assessment within the context of its usefulness. Intellectual assessment is frequently used for evaluation of mental retardation, learning disabilities, traumatic brain injury, psychological disorders including autism and attention-deficit/hyperactivity disorder, and other placement or treatment-related questions (Sattler & Dumont, 2008). In a study of psychologists’ use of intellectual assessment, Harrison and colleagues (1988) identified seven primary purposes of intellectual assessment: to measure capacity, to obtain clinically relevant information, to assess the functional integrity of the brain, to determine educational placement, to determine vocational placement, to develop educational interventions, and to develop vocational interventions. Although Harrison et al.’s (1988) study applied to adult assessment, many of the same purposes apply to assessment of intelligence in children.

Child assessment frequently involves education-related assessment for the purposes of determining whether the child meets eligibility criteria for special education services under the Individuals with Disabilities Education Improvement Act (IDEA, 1990) in one of several categories (Sattler, 2008). These categories include specific learning disabilities, mental retardation, emotional disturbance, other health impairments (including attention-deficit/hyperactivity disorder), autism, traumatic brain injury, and developmental delay, as well
as several others which less frequently utilize intellectual assessment (Sattler, 2008). In addition, child assessment is frequently conducted for private purposes, including treatment placement, treatment planning and social security disability assessment.

Of the varied purposes of intellectual assessment, several purposes involve the use of assessment to make major decisions about educational placement and opportunities. It is of vital importance that assessment be accurate and specific if it is to guide placement decisions, because a poor decision could limit a child’s educational opportunities for several years. Therefore, it is critical that intelligence tests continue to develop into more specific, valid, and reliable instruments with each revision.

With test development, a major concern is the issue of construct validity, or whether the test actually measures what the test developers intend it to measure. In the development of intelligence tests, this occurs at two primary levels. First, the test overall needs to measure general intelligence, often abbreviated as $g$. Second, each subtest must measure not only $g$ but also a more specific cognitive ability related to $g$. One concern with some intelligence test subtests is construct validity of the specific abilities. If a subtest purports to measure verbal knowledge only, for example, evaluators might misinterpret the results of such a subtest if it fact is influenced by fluid reasoning processes. For this reason, subtests are continually scrutinized through research to determine whether they are pure measures of a single ability or whether they are in fact measuring several abilities. Given the interrelatedness of cognitive abilities, it is extremely rare for a subtest to be a “pure” measure of a single ability. Rather, the majority of subtests measure a specific ability along with other abilities required for the successful completion of the test. For example, even the WISC-IV Digit Span subtest, which primarily measures working memory, is influenced by auditory processing and language abilities. This is
not a flaw of this or any other subtest, but it is important when interpreting a test to be aware of the various abilities required to perform well, in order to prevent gross misinterpretation of results.

Construct validity has been established for most of the WISC-IV subtests simply because they are only slightly revised from previous editions of the WISC. In cases of new subtests, however, it is important for researchers to examine the construct validity of the test to determine whether it is a useful measure and what interpretations can be made based upon the results. The purpose of the present study is along this vein. The newest revision of the Wechsler Intelligence Scale for Children (WISC-IV; Wechsler, 2003) included two new subtests: Matrix Reasoning and Picture Concepts. These new subtests are both part of the new WISC-IV Perceptual Reasoning Index and are described as nonverbal measures of fluid intelligence. There is some evidence that these tests also measure verbal ability to a degree, because it is hypothesized by some researchers that one or both of these tasks may involve subvocal verbal reasoning (Dugbartey et al., 1999; Roth, 1978; Sattler & Dumont, 2008). If verbal reasoning is utilized on a measure that is intended to be nonverbal, this may lead to inaccurate interpretation of test results, particularly in cases of language impairment. It is important to determine exactly which skills are assessed by Matrix Reasoning and Picture Concepts so that the tests are utilized accurately and in a manner which can aid interpretation. Therefore, the purpose of the present study is to determine whether verbal mediation is occurring on these subtests in a sample of non-referred children.

If in fact verbal mediation is occurring during completion of the Matrix Reasoning and Picture Concepts subtests, it is expected that the resulting subtest scores will be correlated with test scores representing verbal ability. Specifically, Matrix Reasoning and Picture Concepts
subtest scores will have stronger correlations with the Verbal Comprehension Index (VCI) score than will scores from the third Perceptual Reasoning subtest, Block Design. In addition, it is expected that Block Design scores will be moderately correlated with the Beery-Buktenica Developmental Test of Visual-Motor Integration (VMI 5th Edition) because neither measures verbal ability. Specifically, Block Design will be more closely related to the VMI than will the Matrix Reasoning and Picture Concepts subtests of the WISC-IV.

Unfortunately, although the WISC-IV was published in 2003, there remains a dearth of research regarding the construct validity of the Picture Concepts and Matrix Reasoning subtests. However, it remains true that a body of theoretical and empirical knowledge supports the above hypotheses. When possible, research directly focused on the WISC-IV Picture Concepts (PC) and Matrix Reasoning (MR) subtests will be reviewed. In lieu of studies that directly evaluate the Picture Concepts and Matrix Reasoning subtests, studies that investigate similar nonverbal reasoning subtests or historical precursors to these measures will be discussed.

**Literature Review**

I begin this literature review with a discussion of the construct of intelligence, its definition, and implications.

**What is Intelligence?**

As a science, psychology has long debated the definition of intelligence. It is a debate which has political and emotional implications for most everyone involved. However, in recent years, one theory has been supported by research more than any others, and is now beginning to have a major impact on test development and interpretation. This theory, the Cattell-Horn-Carroll (CHC) theory of intelligence, suggests that although intelligence can be summed up with g, it is more helpful and accurate to look at the seven primary cognitive abilities which comprise
These cognitive abilities are intended to include the construct of intelligence as well as other cognitive abilities which can affect the expression of intelligence. Therefore, assessment of all cognitive abilities is ideal. For a representation of the seven CHC abilities and related subtests of the WISC-IV, see Table 1.

The first such ability in the CHC model is fluid intelligence (Gf), which refers to those reasoning skills utilized on novel tasks which are not yet automatic (Flanagan, Ortiz, & Alfonso, 2007). Fluid intelligence includes concept formation, pattern perception, inferential reasoning, problem-solving, inductive reasoning, and deductive reasoning. On the WISC-IV, fluid intelligence is measured by the Matrix Reasoning, Picture Concepts, and Arithmetic subtests (Wechsler, 2003; Beal, 2004). Of these, Matrix Reasoning and Picture Concepts are identified as specifically measuring a subtype of fluid intelligence, induction. Arithmetic is identified as specifically measuring a subtype of fluid intelligence, quantitative reasoning. Interestingly, Block Design is not included in Flanagan et al.’s (2007) description of fluid intelligence, despite its aspects of pattern recognition, problem solving, and inductive and deductive reasoning. Instead, Flanagan et al. (2007) and others (Sattler & Dumont, 2008) emphasize the visual and spatial processing skills used to complete the Block Design task.

The second major ability according to CHC theory is crystallized intelligence (Gc), which refers to the breadth and depth of verbal knowledge that has been gained through culture, education, and other learning experiences (Flanagan, et al., 2007). On the WISC-IV, crystallized intelligence is measured by the subtests which comprise the Verbal Comprehension Index: Similarities, Vocabulary, Word Reasoning, Comprehension, and Information. Of these, Similarities, Vocabulary, and Word Reasoning are identified as measuring a subtype of
crystallized intelligence called lexical knowledge. Comprehension and Information are identified as measuring a subtype of crystallized intelligence called general information.

The third major ability according to CHC theory is short-term memory ($G_{sm}$), which refers to the ability to hold and utilize information within a few seconds, after which the individual generally either forgets or learns the information. In CHC theory, short-term memory includes the construct of working memory, which involves a greater degree of interaction with and manipulation of the information in short-term memory (Flanagan, et al., 2007). It includes the phonological loop and the visuospatial scratchpad proposed in the cognitive psychology literature (Baddely, 1992). In addition, it includes learning processes and simple memory span. On the WISC-IV, short-term memory is measured by the following subtests: Digit Span, Arithmetic, and Letter-Number Sequencing. Digit Span measures the narrow ability of memory span. Arithmetic and Letter-Number Sequencing measure the narrow ability of working memory.

The fourth major cognitive ability according to the CHC model is visual processing ($G_v$), which refers to the ability to perceive, analyze, organize, and generate patterns and relationships among visual stimuli (Flanagan, et al., 2007). On the WISC-IV, visual processing is measured by the Block Design and Picture Completion subtests. Block design is purported to measure the specific ability of spatial relations, and Picture Completion is described as a test of ‘flexibility of closure’ (Flanagan, et al., 2007).

The fifth major CHC cognitive ability is auditory processing ($G_a$), which refers to the ability to perceive, analyze, organize, and generate patterns and relationships among auditory stimuli as well as discriminate among subtle differences in sound (Flanagan, et al., 2007). On the WISC-IV, there are no subtests of auditory processing. In fact, when psychologists follow the CHC model closely for their assessment of cognitive abilities, they generally turn to a cross-
battery approach and incorporate subtests from the Woodcock-Johnson Tests of Cognitive Abilities or another appropriate test (Flanagan, et al., 2007).

The sixth major cognitive ability according to the CHC model is long-term storage and retrieval ($G_{lr}$), which refers to the processes utilized in learning and encoding of information, storage of information, and later retrieval of the information (Flanagan, et al., 2007). This process is most closely tied to lay concepts of learning and memory. On the WISC-IV, there are no subtests of long-term storage and retrieval. Again, psychologists who closely follow the CHC assessment model supplement the WISC-IV with subtests from another test of intelligence or memory (Flanagan, et al., 2007).

The seventh and final major CHC cognitive ability is processing speed ($G_s$), which refers to the speediness and efficiency with which a person can perceive and respond to information that is either simple or over-learned (Flanagan, et al., 2007). Processing speed is also conceptualized as ‘mental quickness’ which is commonly associated with intelligence. Flanagan et al. describe processing speed as “the ability to fluently and automatically perform cognitive tasks, especially when under pressure to maintain focused attention and concentration” (2007, p. 291). On the WISC-IV, processing speed is measured by the Cancellation, Symbol Search, and Coding subtests. Symbol Search and Cancellation are described as subtests which measure the specific ability of perceptual speed. Coding is described as the subtest which measures the narrow ability of rate-of-test-taking (Flanagan, et al., 2007).

Not every intelligence test is developed to parallel the CHC model. In fact, the WISC-IV was developed both in deference to Wechsler tradition and to the CHC model’s rapidly increasing popularity and empirical support (Flanagan & Kaufman, 2004). In examining the field’s assessment of intelligence or cognitive abilities, it is notable that assessment tools have
only begun to parallel the CHC model in recent years. Initially, it followed a very different model.

**Assessment of Intelligence**

Cognitive assessment has importance for the variety of reasons discussed above, and has therefore been an important focus of psychologists’ efforts. The intellectual testing movement began toward the latter half of the 19\(^{th}\) century when Sir Francis Galton developed sensorimotor tests (Flanagan & Kaufman, 2004) under the hypothesis that testing the senses used to seek knowledge would also test the underlying intelligence. Although this is not intelligence testing as we perceive it today, it was founded in the theory that intelligence is developed through receiving information via the senses; therefore, those individuals with the best sensory development can be expected to be the most intelligent (Flanagan & Kaufman, 2004).

Contemporary intellectual assessment was greatly impacted by Alfred Binet and his colleagues when they compiled a series of tasks into what would eventually become the Stanford-Binet (SB5; Roid, 2003), currently in its fifth edition, for the purpose of differentiating between different levels of cognitive impairment or mental retardation (Kaufman & Lichtenberger, 2006). The next major revision to the purpose and process of intellectual testing involved the Army Alpha and Army Beta tests which were developed by the United States during World War II for the purpose of predicting which recruits had the aptitude for officer training (Wechsler, 2003; Kaufman & Lichtenberger, 2006). These tests added nonverbal measures to cognitive assessment. The Army Alpha test was designed to measure verbal reasoning abilities. The Army Beta test was considered a performance test and was intended for non-English-speaking and other nonverbal individuals, with the intention of measuring nonverbal reasoning skills (Kaufman & Lichtenberger, 2006). The goal of this second measure
was to differentiate between those individuals who actually had low ability and those who were malingering to avoid service.

The next major change to cognitive assessment occurred when David Wechsler, a psychologist at New York’s Bellevue Psychiatric Hospital, began to see the clinical usefulness of cognitive assessment. Prior to Wechsler’s work, assessment of children was utilized primarily for school placement and diagnosis of cognitive disabilities, and assessment of adults was utilized primarily for placement within the United States Army. Wechsler argued that intellectual assessment had more wide-range clinical utility, and developed a test which incorporated subtests of established intelligence measures for clinical applications (Flanagan & Kaufman, 2004). For this reason, he was the first to combine verbal and performance measures, and to take the somewhat radical step of giving performance measures equal weight as verbal measures (Flanagan & Kaufman, 2004; Kaufman & Lichtenberger, 1999, 2006). His first intelligence test, published in 1939, was titled the Wechsler-Bellevue Intelligence Scale (Flanagan & Kaufman, 2004). Wechsler quickly saw the usefulness of having both adult and child forms of his intelligence test, and developed the Wechsler-Bellevue Intelligence Scale – Form II in 1946 to address child assessment (Flanagan & Kaufman, 2004). Although the Wechsler measures have been revised several times, both the adult and child versions followed a similar format for decades, and rapidly became the gold-standard measure of child and adult assessment. Even with the incorporation of CHC theory into recent test development, the Wechsler scales have adapted and retained their authority over the field of intellectual assessment.

**Evolution of the WISC**

The WISC-IV is a direct descendant of the Wechsler-Bellevue Form II (WB-Form II). Beginning with the WB-Form II, the field of child intellectual assessment began to flourish.
Although the WB-Form II was not exclusively developed for children, its age range extended from age 10 to age 79 and it was the precursor of the WISC scales (Flanagan & Kaufman, 2004). This represented the beginning of utilizing child intellectual assessment outside of identification of mental retardation and learning disabilities (Flanagan & Kaufman, 2004). Intellectual assessment of children was now utilized for clinical assessment, including diagnosis and treatment planning. In addition, tests of intelligence became more reliable and wide-ranging with the WB-Form II, marking a departure from a period in which intelligence was determined either through single tasks or multiple tasks of a single ability (Kaufman & Lichtenberger, 2006). As an example of the increasing breadth, the WB-Form II gave equal weight to verbal and performance measures, rather than relying exclusively on tests of verbal ability.

The WB-Form II was revised in 1949 and became the first Wechsler test specifically developed for children. This original WISC was developed for children aged 5 to 15, and was intended to parallel the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955). As the field of child assessment continued to rapidly evolve, Wechsler developed and published an updated revision of the WISC in 1974 (WISC-R; Wechsler, 1974). This revision was intended to update the normative sample, to update the materials for improved durability and child-friendliness, and clarify the standardization of administration and scoring procedures (Flanagan & Kaufman, 2004).

The next major revision to the WISC was in 1991, with the publication of the WISC-III (Wechsler, 1991). The WISC-III continued to give equal weight to performance and verbal ability by dividing the full-scale intelligence quotient (FSIQ) into the verbal and performance intelligence quotients (VIQ and PIQ, respectively). However, two new index scores were added to enhance clinical utility, the Freedom from Distractibility Index (FDI) and the Processing
Speed Index (PSI; Flanagan & Kaufman, 2004). The FDI was intended to provide information about attentional difficulties, and to determine how impacted the other cognitive scores were by the child’s ability to attend to the test items. The PSI was intended to provide information about how speed of performance may have impacted other scores separate from verbal and performance abilities. The structure of the test now had the VIQ divided into the FDI and the Verbal Comprehension Index and the PIQ divided into the PSI and the Perceptual Organizational Index. These revisions were important for maintaining the WISC as a strong clinical measure, but were insufficient to keep the Wechsler scales current with CHC theory, which was beginning to rise in importance (Flanagan & Kaufman, 2004).

The latest WISC revision was published in 2003, a mere 12 years after its most recent predecessor (Wechsler, 2003). The test was restructured, and although Wechsler did not report specifically following CHC theory, research regarding CHC theory was cited extensively in the test manual (Flanagan, et al., 2007; Wechsler, 2003) and CHC theory is reflected to a greater degree than on any previous WISC (Kaufman et al., 2006). The revision was arguably the most extensive revision on a Wechsler test since the original Wechsler-Bellevue. See Table 2 for the organization of WISC-IV subtests and indexes.

In this fourth revision, the VIQ and PIQ were eliminated from the WISC-IV. Instead, the FSIQ is subdivided into four index scores, which are similar but not identical to the old index scores (Beal, 2004; Sattler & Dumont, 2008; Wechsler, 2003). The Perceptual Reasoning Index (PRI) is similar to the old Perceptual Organizational Index (POI), but with increased emphasis on fluid reasoning and decreased emphasis on visual processing ability. Three of the old POI subtests were eliminated from the WISC-IV because they relied too heavily on fine motor movements, and it was often difficult to determine whether low scores were related to low
intelligence or motor impairment. These three subtests have been replaced by the Matrix Reasoning subtest, which was first introduced on the WAIS-III, and the Picture Concepts subtest, which is entirely new (Sattler & Dumont, 2008). The Verbal Comprehension Index (VCI) is similar to the old VCI. However, one of the old core subtests, Information, became supplemental, and Word Reasoning was introduced as a supplemental subtest (Sattler & Dumont; Wechsler, 2003). The Working Memory Index (WMI) is based upon the old Freedom from Distractibility Index, and retains Digit Span as one of the core subtests. However, Arithmetic became a supplemental subtest, and Letter-Number Sequencing, originally developed for the WAIS-III, was introduced as the second core WMI subtest (Sattler & Dumont; Wechsler). The Processing Speed Index (PSI) was retained, as were the Coding and Symbol Search subtests. Cancellation was introduced as a supplemental test for the PSI (Sattler & Dumont).

In addition to these major structural revisions, the WISC-IV also includes a new normative sample, which is reported to be more inclusive than the old sample (Wechsler, 2003). In addition, graphics were again revised to remain updated and child-friendly. This most recent WISC has a total of 15 subtests, 10 of which are core.

The normative sample for the WISC-IV included 2,200 children stratified on the basis of age, sex, race/ethnicity, geographic region, and parental education to match 2000 U.S. census data (Wechsler, 2003). The children were divided into 11 age groups, with 100 boys and 100 girls in each age group. The Arithmetic subtest was an exception, and was standardized on exactly half of the total sample. According to Sattler and Dumont, the “sampling methodology is considered to be excellent” (2008, p. 267).

The WISC-IV is also well-regarded for its psychometrics. According to the WISC-IV statistics reported in the manual (Wechsler, 2003), the average internal consistency reliabilities
of subtests ranges from a low of .68 (Symbol Search and Cancellation-Random) to a high of .89 (Vocabulary, Matrix Reasoning, and Letter-Number Sequencing). The internal consistency reliability of the FSIQ itself is .97. The average test-retest reliabilities for the subtests range from a low of .67 (Digit Span-Backward) to a high of .85 (Vocabulary). The test-retest reliability for the FSIQ is .89. Overall, Sattler and Dumont describe the reliability of the WISC-IV as “outstanding” (2008, p. 270).

In terms of validity, the Wechsler scales are adequate. Criterion validity has been assessed by comparing WISC-IV scores to scores on the WISC-III, WPPSI-III (Wechsler Preschool and Primary Scale of Intelligence-III), and the WAIS-III. The WISC-IV scores were moderately to highly correlated with the WISC-III scores (.85 for the VCI and .70 for the POI/PRI), which is unsurprising given that 70% of WISC-IV items can be found on the WISC-III (Sattler & Dumont, 2008). The WISC-IV scores were also moderately to highly correlated with the WPPSI-III scores (.76 for verbal scales, .74 for the PRI/performance scale, and .85 for the FSIQ). Finally, WISC-IV scores were also moderately to highly correlated (.84 for verbal scales, .71 for the PRI/performance scale, and .88 for the FSIQ) to the WAIS. It would be useful to know more about the criterion validity of the WISC-IV as compared to a non-Wechsler test of intellectual ability, such as the Stanford-Binet 5th Edition, but all research was done within the WISC publisher, Harcourt Assessment (Sattler & Dumont, 2008). That the quality of the psychometrics has been evaluated primarily by the publisher of the test represents a weakness given that the test publishers are likely to experience some degree of conflict of interest or bias in evaluating their own product.

Construct validity was assessed by comparing correlation coefficients between subtest scores and index scores. Verbal comprehension subtests correlated with the VCI from .70 to .91.
Perceptual reasoning subtests correlated with the PRI from .57 to .84. Working memory subtests correlated with the WMI from .57 to .86. Finally, processing speed subtests correlated with the PSI from .41 (Cancellation) to .88 (Sattler & Dumont, 2008). Although most of these correlation coefficients are moderately high, the subtest scores are being correlated with the indexes of which they comprise a portion of the score, which results in a certain degree of item overlap, thereby reducing the confidence one can have in these correlation coefficients.

Sattler and Dumont (2008) conducted a factor analysis of the WISC-IV subtests to determine whether the four index structure was supported. Not only was the format of the indexes supported, but Sattler and Dumont found that the construct validity coefficients reported in the WISC-IV manual had a strong positive relationship with the $g$ factor ($\rho = .91$, $p < .01$).

Overall, the WISC-IV has excellent psychometric properties and a representative normative sample. In addition, it has followed research and theory to produce a revision which is compatible with current models of intellectual assessment. Therefore, it can be described as an outstanding test overall, although the new subtests should be the subject of further research. Two subtests that are new to the WISC-IV are discussed in greater detail below as they are the foci of the present study. The Matrix Reasoning subtest was first introduced on the WAIS-III, but the Picture Concepts subtest is entirely new to the Wechsler series.

**Matrix Reasoning Subtest**

The Matrix Reasoning subtest is one of three core subtests on the Perceptual Reasoning Index of the WISC-IV. This subtest has 35 items and is not timed. The child is presented with a series of incomplete matrices, each of which is a series of abstract patterns and designs. The child is directed to select the best from among several answer choices in order to complete the matrix (Sattler & Dumont, 2008). The matrices are thought to be comprised of four general
types, including continuous and discrete pattern completion, classification, analogical reasoning, and serial reasoning (Gabel, 2003; Williams, Weiss, & Rolfus, 2003a). For an example, see Figure 1.

The Matrix Reasoning subtest was originally adapted from Raven’s Progressive Matrices (Raven, 1965) for the Wechsler Adult Intelligence Scale (Williams, et al., 2003a). On the newest WISC-IV, Matrix Reasoning, along with Picture Concepts, has replaced the Object Assembly and Picture Arrangement subtests as a measure of fluid intelligence which is less confounded by fine motor skills. Although Matrix Reasoning is ostensibly a measure of fluid intelligence (Gabel, 2003), there is some evidence that Matrix Reasoning also involves visuo-sensory construction skills and verbal mediation skills (Sattler & Dumont, 2008). Although examinees are not asked to state their reasoning and thinking aloud on this subtest, it is expected that verbal mediation is occurring during the process of pattern recognition and inductive reasoning.

The psychometrics of the Matrix Reasoning test included in the WISC-IV are generally considered to be quite strong. According to the WISC-IV Technical Manual (Wechsler, 2003), Matrix Reasoning has an excellent average test-retest reliability of $r = .85$ and an average internal consistency reliability of $r = .89$. It shares a moderate correlation with the Picture Concepts ($r = .47$) and Block Design ($r = .55$) subtests, which together comprise the Perceptual Reasoning Index. It appears to have a moderate to high correlation with the Perceptual Reasoning Index itself ($r = .64$), and only a moderate correlation with the Verbal Comprehension Index ($r = .52$; Wechsler, 2003). Overall, these psychometrics can be described as excellent, and are a primary reason the Wechsler scales have become the gold standard for assessment of cognitive ability.
**Picture Concepts Subtest**

The Picture Concepts subtest is one of three core subtests on the Perceptual Reasoning Index of the WISC-IV (Sattler & Dumont, 2008). This subtest has 28 items and is not timed. The child is presented with two or three rows of pictures and is directed to select one picture from each row, such that the selected pictures selected relate to one another conceptually. The subtest begins with items which have only two rows and for which the category is a concrete one. Toward the end of the test, the child is presented with three rows of pictures and the categories become abstract (Gabel, 2003). For example, the child might be expected to select two objects at the beginning of the test because both are fruit, whereas they might be expected to select three objects toward the end of the test because each of the three objects uses air to function. This developmental progression is intended to differentiate between children across age groups. For an example item, see Figure 2.

The Picture Concepts subtest was not developed from any existing test (Flanagan & Kaufman, 2004). Therefore, the use of this subtest needs to be supported through further validation studies. According to the WISC-IV manual, Picture Concepts has more than adequate reliability with an average test-retest reliability of $r = .71$ and an average internal consistency reliability of $r = .83$ (Wechsler, 2003). It also appears to be a valid measure of fluid reasoning ability, as it correlates more highly with Matrix Reasoning ($r = .47$) than any other WISC-IV subtest (Sattler & Dumont, 2008). In addition, it has a moderate to high correlation with the Perceptual Reasoning Index ($r = .77$) and only a moderate correlation with the Verbal Comprehension Index ($r = .47$), which supports that this subtest is a better measure of those abilities measured by the Perceptual Reasoning Index than it is a measure of those verbal abilities measured by the Verbal Comprehension Index (Sattler & Dumont, 2008). Unfortunately,
no further information is available at this time about the validity and specificity of the Picture Concepts subtest of the WISC-IV, because few studies of the Picture Concepts subtest have been published to date.

A relatively recent dissertation represents the single published study which specifically addresses the Picture Concepts subtest. Kain (2006) administered the Picture Concepts subtest, as well as several other verbal ability and categorical reasoning subtests, to a sample of approximately 50 fifth and sixth grade students. Kain then deviated from standard administration to ask students to explain their reasoning processes behind each response so that reasoning style and quality could be rated. Kain determined that verbal ability did not appear to be significantly related to scores on the Picture Concepts subtest. Instead, he found that students utilized both thematic- and similarity-based reasoning to arrive at answers. In fact, approximately one-fourth of incorrect responses were identified as quality responses based upon the reasoning utilized, and many correct responses were identified as poor quality responses. As this was the first study specifically examining the reasoning styles utilized on this subtest, more information and replication is needed.

**Verbal Mediation**

A criticism within the literature of the Picture Concepts and Matrix Reasoning subtests relates to whether or not purely visual and reasoning abilities are used to complete these subtests. Instead, it is hypothesized that verbal mediation is also involved, and that incorporation of this strategy may affect clinicians’ interpretation of resulting scores.

Verbal mediation is a complex construct. Verbal mediation strategies are considered to be useful cognitive problem-solving strategies. However, verbal mediation is a developmental process, and begins in early childhood with thinking aloud in what appears to be egocentric
speech (Wertsch, 1985). During early childhood, this egocentric speech is not always utilized for reasoning. Instead, young children often narrate their own play and behavior out loud. According to Vygotsky’s theory of language development, a normally developing child eventually learns to internalize this thought process, and although the speech will no longer be heard by an observer it will continue to influence behavior, reasoning, and planning (Wertsch, 1985). In fact, Vygotsky believed that the primary goal of internalized speech is “communication with the self for the purpose of self-regulation, or guiding one’s own thought processes and actions” (Berk & Winsler, 1995, p. 37). In essence, verbal mediation involves the use of thought processes or internal dialogue to guide problem solving or goal-directed behavior. Because it is a highly useful but unobservable skill, it is difficult to determine when a child might be engaging in verbal mediation strategies after approximately age seven, when this process becomes internalized (Wertsch, 1985).

Verbal mediation strategies are of importance for several reasons. First, they have been found to be an effective problem-solving strategy, and it is possible that children who use these strategies will actually display superior performance on some complex tasks (Bivens & Berk, 1990; Neuman, Leibowitz, & Schwartz, 2000; Ostad & Sorensen, 2007). Second, if verbal mediation strategies are utilized on an assessment measure which is purported to measure only abstract, nonverbal reasoning, the validity of the test must be called into question. Third, it is possible that on the same abstract, nonverbal task, children with language impairments may perform less well than would be expected because of their difficulty effectively applying verbal mediation strategies.

Another concern about the use of verbal mediation on a test of visual and reasoning ability relates to the fact that verbal mediation is impaired in some children who may display
otherwise normal cognitive profiles. For example, utilization of verbal reasoning skills is a common target of attention-deficit/hyperactivity disorder and externalizing disorder treatment, either directly or indirectly (Kazdin, 2003; Webster-Stratton & Reid, 2003). Furthermore, use of verbal skills and more general forms of language and communication is frequently impaired in individuals with autistic disorder and other pervasive developmental disorders (Lovaas & Smith, 2003).

**Verbal Mediation Research**

In what was arguably a classic study of verbal mediation, Kohlberg, Yaeger, and Hjertholm (1968) observed verbal mediation across a variety of ages and tasks of varying difficulty. The authors found a curvilinear relationship between verbal mediation strategies and age, such that children initially increased their use of audible ‘egocentric speech’ with age until approximately age 5 to 6. At this point, children began to reduce their use of egocentric speech, which is consistent with Vygotsky’s theory that such speech becomes internalized and therefore inaudible (Kohlberg, et al., 1968). Interestingly, the authors also observed that the use of egocentric speech increased for difficult tasks compared to simple tasks (Kohlberg, et al., 1968) which is consistent either with children using the strategy more frequently for difficult tasks or resorting to an earlier developmental skill-set for difficult tasks.

As a problem-solving strategy, there is abundant evidence that verbal mediation is helpful. To examine verbal mediation, researchers frequently observe the private but audible speech of children who are too young to have internalized this speech process, generally referred to as private speech. One such study, conducted by Bivens and Berk (1990), followed students longitudinally from first to third grade, and involved classroom observations with specific reference to students’ academic performance, behavior, and use of private speech. The
researchers found that children who engaged in higher frequencies of private speech also demonstrated superior mathematics performance and improved attendance to their work.

In a similar study, Ostad and Sorensen (2007) observed children between the 4th and 7th grades during mathematics. Comparisons were made between those children who were achieving adequately in mathematics and those who demonstrated significant difficulty with math. Ostad and Sorensen noted that average-achieving students demonstrated the expected developmental progression of audible private speech to inaudible private speech (i.e. mumbling or speech too low in volume to be overheard by a researcher) and eventually to fully internalized speech. The students with mathematics difficulty were less likely to reach the internalized speech phase, and continued to utilize inaudible private speech later than would have been developmentally expected (Ostad & Sorensen, 2007). Although it is difficult from this particular study to determine whether the mathematics difficulty necessitated the continued use of a more basic form of verbal reasoning or whether poor use of verbal reasoning led to difficulty understanding the mathematics concepts, it appears that verbal mediation is connected with ability to achieve in mathematics.

Neuman and colleagues (2000) designed a study to discover subtypes of verbal mediation, and to determine whether these types are universally helpful or variable. Participants were 9th grade students who were asked to verbalize their ‘self-explanations,’ or think aloud, while engaging in creative nonverbal problem-solving tasks. The researchers identified five subtypes of verbal mediation: clarification, inference, justification, monitoring, and regulation. In the study, use of clarification and inference self-explanations predicted successful and speedy problem-solving (Neuman, et al., 2000). This suggests that although verbal mediation is helpful, there are particular types which are most predictive of success at certain tasks.
Verbal Mediation and Nonverbal Fluid Reasoning Tasks

If verbal mediation strategies are utilized during completion of the Picture Concepts and/or Matrix Reasoning subtests of the WISC-IV, this would suggest that neither subtest is a pure measure of fluid reasoning and nonverbal analytical ability. Rather, this would suggest that language is also utilized to arrive at solutions to the problems. In fact, there is already some speculation among psychologists that verbal mediation strategies are utilized for Matrix Reasoning tasks (Sattler & Dumont, 2008). In addition, clinicians utilizing CHC cross-battery assessment recognize that Picture Concepts has a moderate degree of linguistic demand and that it is related to crystallized intelligence, which includes verbal knowledge and ability, at certain ages (Flanagan et al., 2007). This notion is supported by Sattler, Dumont, and Rapport (2008), who argued that, among other abilities, the Picture Concepts subtest taps into crystallized knowledge (which is generally acquired verbally through formal education), expressive language, and reading patterns. They did not make the same arguments for the Matrix Reasoning test (Sattler, Dumont, & Rapport, 2008).

Silverberg and Buchanan (2005) studied the impact of verbal mediation on visual memory tasks. After determining which of a variety of visual designs were easiest and most difficult to describe verbally, they engaged adult participants in a visual memory task. Visual designs were shown in series and then participants were later asked to identify which designs had been presented earlier. Some participants were asked to concurrently perform a verbal interference task, some were asked to concurrently perform a visual interference task, and a control group was simply allowed to respond to the visual memory task. Silverberg and Buchanan (2005) found that for items that were easy to describe verbally, the verbal interference group had significantly reduced performance compared to the visual interference and control
groups. Therefore, although visual memory is typically considered to be distinct from verbal ability, it is apparent that verbal mediation is utilized in many instances and improves performance.

In a study of verbal mediation on a matrix task which emphasized inductive reasoning, Welsh (1987) attempted to demonstrate the existence of verbal mediation on such tasks by interfering with the process. She studied children who were categorized either as impulsive or reflective and had the children simultaneously complete a matrix reasoning and simple verbal task. Welsh hypothesized that both groups of children would exhibit impairment on the task due to disrupted verbal mediation, but that the pattern would be more pronounced in the impulsive group. Although she did find that verbal mediation had been disrupted for both groups, the reflective group was actually more impaired by this interference. Welsh hypothesized that the impulsive children may have been using a visual strategy in place of verbal mediation, which was not disrupted by her verbal task. Despite this interesting and unexpected finding, Welsh successfully demonstrated that verbal mediation is vital to success on an inductive reasoning matrix task, particularly for certain types of children.

In a study of Raven’s Advanced Progressive Matrices, DeShon, Chan, and Weissbein (1995) required participants to concurrently verbalize their thought processes and complete the matrices. The authors found mixed results, where some matrices consistently seemed more difficult for participants verbalizing their thought process and other matrices consistently seemed easier for participants verbalizing their thought processes. The authors suggested this may mean that some matrices are more easily verbally mediated, whereas others are more easily visually mediated. The authors did not identify any noticeable differences between those items which were more easily verbally mediated and those which were not, but it is possible that a slightly
Evidence for verbal mediation on supposedly nonverbal fluid reasoning tasks is further supported by a study of the Matrix Reasoning subtest of the WAIS-III (Dugbartey et al., 1999), which is highly similar to the Matrix Reasoning subtest of the WISC-IV. In this study, the researchers compared scores on the WAIS-III Matrix Reasoning test to scores on a test of verbal abstract reasoning both for English-speaking United States-native adult participants and for non-English-speaking immigrant participants. The researchers found that success on the Matrix Reasoning subtest was correlated with left-hemisphere dominance in the brain and with the utilization of verbal mediation strategies (Dugbartey et al., 1999), and that this pattern was most pronounced with the non-English-speaking participants. Importantly, left-hemisphere dominance is typically associated with a pattern of higher verbal abilities than nonverbal abstract reasoning abilities. The authors noted that, given the importance of the left-hemisphere to success at the task, the Matrix Reasoning task is not truly nonverbal. In fact, the authors of this study went so far as to describe the description of Matrix Reasoning as a nonverbal task “misleading,” and indicated a superior term might be “non-enuntiary,” meaning not enunciated out loud (Dugbartey et al., 1999, p. 400). This suggests that the description of this task as nonverbal, simply because it does not require vocalizations from the examinee, is likely inaccurate, given that verbal abilities appear to contribute to success on the test.

Although the above cited studies apply most directly to the Matrix Reasoning subtest, they also apply indirectly to the Picture Concepts subtest because both are intended to measure the same ability. However, there is also some separate evidence for the role of different forms of reasoning would have been applicable to the disparate types of matrices (DeShon et al., 1995).
reasoning on the Picture Concepts test, including thematic-based and similarity-based reasoning styles, although these may not represent forms of verbal reasoning per se (Kain, 2006).

Unfortunately, there is a complete lack of research that specifically addresses the Matrix Reasoning and Picture Concepts subtests on the WISC-IV, particularly with regard to the role verbal mediation skills may play in completing the task. Although the WISC-IV was first published in 2003, it remains true 8 years later, our understanding of the abilities used to successfully complete the Matrix Reasoning and Picture Concepts tasks remains incomplete.

**Rationale for Study**

There is evidence that verbal mediation may take place on some tasks that have traditionally been considered to be nonverbal reasoning tasks. However, little research has been done with regard to matrix tasks, particularly on intelligence tests for children, and none has been conducted with regard to picture concepts-type tasks. Therefore, few conclusions can be drawn about the role of verbal mediation on the completion of these two WISC-IV subtests. Although the WISC-IV was published approximately eight years ago, the gap in research remains. Unfortunately, this means that psychologists are using two new subtests as pure measures of fluid reasoning without any independent validation that these subtests are measuring fluid reasoning exclusively. In fact, there is some evidence that these tests are impure measures, and that verbal mediation abilities may be confounding results. Particularly in cases of speech and language impairment or other significant variability in cognitive abilities, such as is typical with some diagnoses, including ADHD and Autism-spectrum disorders, this may lead to poor decision making about some children based upon test results.

This study is intended to begin to fill the gap in research by assessing for the presence of verbal mediation in a non-clinical sample of children on Matrix Reasoning and Picture Concepts
subtests. They will be compared to the third Perceptual Reasoning Index subtest, Block Design. In contrast to the Matrix Reasoning and Picture Concepts subtests, the Block Design subtest is reported to have increased reliance on visuo-spatial skills. The goal of the present study is to investigate the relationship between these measures of fluid reasoning ability and established measures of verbal ability and visuo-spatial skills in a community sample of children and to further elucidate the skills utilized in the completion of these subtests, so that scores can be interpreted accurately. For the purposes of this study, the Similarities subtest from the WISC-IV has been selected as a verbal measure. The Similarities subtest is one of three core subtests in the Verbal Comprehension Index, and is identified as the verbal measure most heavily relying upon fluid reasoning ability. The Beery-Buktenika test of Visual Motor Integration, or VMI, has been selected as the measure of visuo-spatial skills. Correlations will be obtained between each subtest from the Perceptual Reasoning Index and the two aforementioned contrast measures to determine whether the target subtests are more reliant upon verbal abilities or visuo-spatial skills. It has already been firmly established that the three subtests utilize fluid reasoning skills.

The following hypotheses will be investigated in this study:

**Hypotheses for Matrix Reasoning**

1. It is hypothesized that Matrix Reasoning will have a higher correlation with Picture Concepts than with Block Design.

2. It is hypothesized that Matrix Reasoning will have a higher correlation with Similarities relative to the correlation between Block Design and Similarities.

3. It is hypothesized that Matrix Reasoning will have a lower correlation with Beery’s VMI than will Block Design and Beery’s VMI.
Hypotheses for Picture Concepts

4. It is hypothesized that Picture Concepts will have a higher correlation with Matrix Reasoning than with Block Design.

5. It is hypothesized that Picture Concepts will have a higher correlation with the Similarities subtest relative to the correlation between Block Design and Similarities.

6. It is hypothesized that Picture Concepts will have a lower correlation with Beery’s VMI than will Block Design and Beery’s VMI.

Method

Participants

Participants in this study include a sample of children aged 6-17 who were recruited from the general community by students in a graduate psychology program. Participants primarily resided in the Pacific Northwest, but were also recruited from additional locations in the United States and Canada. Data was analyzed from a total of 76 participants, with ages varying from 6 years to 16 years. The average age of the sample was 10 years and 8 months, with a standard deviation of about 3 years ($SD = 3.22$). Approximately sixty-three percent of the participants were female. Participants were asked to report their race and ethnicity for the study. The reported racial and ethnic makeup of the sample is approximately 82 percent Caucasian, 9 percent biracial, 4 percent Asian American (including Vietnamese American, Chinese American, and Indian American), 3 percent Hispanic/Latino, and 3 percent of participants chose not to report their racial and ethnic background. The average IQ score for the sample was 111 ($SD = 11$), with scores ranging from 81 to 132. See Table 3 for a further summary of subtest score distributions.
Procedure

All tests were administered by graduate students in a professional psychology program who were currently or previously enrolled in a child assessment course. Because tests were administered over the course of several years after the publication of the WISC-IV, a total of 71 test administrators were included in the study. Administrators were in their second year of the psychology program at a minimum, and had previous training in administration of intellectual and personality tests to adults, psychometrics, ethics, clinical skills, and diagnostic interviewing. Furthermore, administrators were required to pass a competency check on the WISC-IV with a more senior graduate student prior to administering tests. Specifically, the competency check involved administration of the WISC-IV and other measures to a senior graduate student functioning as a teaching assistant for the course. Competency checks were designed to be standardized and to elicit understanding of subtest-specific prompts, discontinue rules, and other potentially challenging standardization procedures. All administrators passed competency checks on their first attempts; however, the participant protocols included in this study were reviewed for accuracy regardless, creating a further check for accuracy and standardization of administration.

Participants were recruited by trained graduate students, and tended to be in acquaintance-level relationships with the administrators. To control for any potential bias or errors in scoring due to the fact that the administrators were still learning the test, scoring of tests was reviewed by a more senior graduate student prior to inclusion in the database. Administration of the tests was not observed or videotaped, and there is some possibility that standardization procedures were not followed equally by all administrators. For example, although administrators were directed to follow standard administration as described in the
WISC-IV manual, administration was completed in a variety of locations, and factors such as lighting, table space, and seating may have varied somewhat. Although this is a potential weakness of the study in terms of standard measurement, it also reflects actual clinical practice, and may improve the external validity of the study.

Prior to administration of the tests, written informed consent was obtained from parents and guardians of all participants, and written assent was obtained from the children themselves. Participants were generally offered a small monetary incentive for their participation in the study, and no consequences existed if participants chose to withdraw.

Tests were administered between 2003 and the present. All participants were administered the 1) Wechsler Intelligence Scale for Children-IV, 2) the Beery Developmental Test of Visual Motor Integration, Fifth Edition (VMI-5th Edition), and 3) selected subtests of the Wechsler Individual Achievement Test. Completion of all three portions of testing generally took about 3 ½ to 4 hours of participant testing time. For the purposes of the present study, only data from the WISC-IV and the VMI were included although all were administered.

After administration, participants would have been excluded from the statistical analyses if it was determined that a subtest had been administered in a non-standardized manner, if the participant had been recently tested using the WISC-IV, or if relevant data was missing or incorrect. Although administration was not observed directly, evidence of non-standardized administration would be gained if an administrator discontinued a subtest after too few or too many items, began administration at an inappropriate starting point, and the like. Parents of participants were asked directly to report any recent testing. Individuals who had been tested once for the present study were excluded from a second participation in the study. It was determined that relevant data was missing or incorrect if actual testing data was missing, if
scoring was inaccurate, or if calculations appeared to be incorrect. Missing data was not considered to be relevant if it was merely participant name, examiner name, and other information that was deliberately missing for the purpose of confidentiality. Other than the described exclusion criteria, this study was designed to be inclusive of normal variability in a community sample. Exclusion criteria were selected using the rationale that error based upon administration or identified disability should be excluded for this analysis, but that otherwise the sample should be representative of the population and its diversity.

Measures

**Wechsler Intelligence Scale for Children 4th Edition.**

As described above, the WISC-IV (Wechsler, 2003) is an intelligence test for children aged 6 through 16. This test is administered individually following a standardized testing procedure outlined in the test manual. The WISC-IV includes 15 subtests, 10 of which are considered to be core subtests, and 5 of which are considered to be supplemental. Scores are reported in terms of overall cognitive ability (g), which is represented by the FSIQ, as well as four sub-abilities, which are represented by the index scores. These include the Working Memory Index, the Processing Speed Index, the Perceptual Reasoning Index, and the Verbal Comprehension Index (VCI). For the purposes of the present study, the Similarities score, Perceptual Reasoning Index score, and scores on the three Perceptual Reasoning subtests (Block Design, Picture Concepts, and Matrix Reasoning) were used.

**Beery Developmental Test of Visual-Motor Integration 5th Edition (VMI-V).**

The VMI (Beery & Beery, 2004) tests accuracy of copying geometric figures in children aged 2 through 18. This test is administered individually, and is not timed. The VMI items become developmentally more difficult across its 24 items. The VMI specifically measures
graphomotor ability, and can be impacted by a person’s fine motor skills as it involves drawing of increasingly complex figures (Fischer & Loring, 2004).

The VMI has moderate to high reliability coefficients (Graham, McKnight, & Chandler, 2009). Specifically, the VMI manual (Beery & Beery, 2004) reports internal consistency reliability coefficients for children ranging from a low of .79 (for children age 9) to a high of .89 (for children age 3). This suggests individual VMI items are highly consistent in measuring the same underlying skillset. Additionally, the VMI is not highly subjective. The scoring criteria is adequately specific to result in an interscorer reliability of .92 (Beery & Beery, 2004).

In terms of concurrent validity, the VMI has correlations of .66 with the WISC-R, which suggests the VMI adequately measures aspects of g. A moderate correlation is adequate given the low amount of item and content overlap between the two tests (Graham et al., 2009). In terms of construct validity, the test developers examined how strongly scores correlate to age of examinee, because the VMI is intended to measure an ability which progresses developmentally. For the total norming sample, which includes children and adults, age was correlated with raw score on the VMI at .89 (Beery & Beery, 2004). Unfortunately, the test developers did not report child-specific data with regard to construct validity.

Results

Preliminary Analyses

Key subtest scores, including Similarities, Picture Concepts, Matrix Reasoning, Block Design, and Beery’s VMI were assessed for outliers, as well as normality, including an examination of skewness and kurtosis statistics.

Given the goal of being inclusive of normal variability for the community, participant data would only be excluded from the present study if it markedly impacted the resulting
analyses due to outlier scores. Using a cut-off score of 3 distributions above or below the mean identified two scores as outliers. One participant obtained a scaled score of 2 on the Similarities subtest. A separate participant obtained a scaled score of 3 on the Matrix Reasoning subtest. It was determined to include data from the aforementioned participants based on the rationale that it is most inclusive or normal population variability, and because the two scores came from separate participants and therefore do not appear to represent a learning disability or other confounding variable.

Analysis of the skewness and kurtosis statistics suggested that some of the variables experienced a negative skew due to one participant obtaining considerably lower scores. Compared to the normative sample, however, this participant would not have represented an extreme case with an FSIQ of 80. As this score represents normal variability in scores and is a frequently occurring score in the community, the participant was not removed from the sample. Skewness and kurtosis statistics are reported for each subtest score and summary score in Table 4.

In order to assess for the assumption of linearity inherent in correlation analyses, scatterplots were created using SPSS software, and were visually inspected for linearity of distribution. All subtest variables appeared to have adequate linearity.

**Statistical Analyses**

The Similarities subtest was rationally selected as the verbal measure for the following analyses due to consensus among the psychological community that, of the VCI subtests, the Similarities subtest is most closely related to PRI subtests through its reliance on logical reasoning skills (Sattler, 2008). The VCI test which most closely relates to the PRI subtests was selected based upon the rationale that this would reduce risk of Type I error.
Pearson correlation coefficients were computed between the Similarities subtest and each of the three PRI subtests, as well as between the VMI and each of the three PRI subtests. In order to control for Type I error across the multiple tests, the Bonferroni correction was applied such that a $p$ value of .0055 (.05/9) was required for the initial correlation analyses, and a $p$ value of .0083 (.05/6) for the later comparisons. Magnitude of correlations were then compared utilizing a method introduced by Blalock (1972) which statistically evaluates whether significant differences exist between two correlations obtained from a single sample. Blalock’s technique involves applying the following equation to the correlation coefficients from the subtest being compared:

$$t = (r_{xy} - r_{zy}) \sqrt{\frac{(n-3)(1+r_{xz})}{2[(1-r_{xy}^2)(1-r_{zy}^2)+(2r_{xy})(r_{xz})(r_{zy})]}}$$

In effect, this equation compares the two correlation coefficients to one another while accounting for shared variability due to measuring overlapping constructs and due to being derived from the same sample of participants. Additionally, this method takes into account the total sample size. The resulting score is then examined in a t-distribution table as would a traditional t-score. For the current analysis, n-3 resulted in 73 degrees of freedom. As mentioned above, a $p$ value of .0083 was used for the analysis. The hypotheses would be supported if the t-value obtained from Blalock’s calculation is higher than the observed t-value from the t-distribution table, and would not be supported if the value in the table was higher than the t-value obtained from the calculation:

With the exception of the comparison analysis (Blalock, 1972) which was calculated in an excel spreadsheet, and the power analyses, which were calculated using G*Power (Faul, Erdfelder Buchner, & Lang, 2009), all analyses were computed using the Statistical Package for the Social Sciences (SPSS, version 18.0) software.
The selection of the Similarities subtest as the verbal measure for later correlation analyses was evaluated and confirmed statistically using a multiple regression analysis with each of the three VCI subtests and the VCI itself as predictor variables, and each of the three PRI subtests as criterion variables. The results of this analysis indicated the Similarities subtest was the best predictor of scores on the various PRI subtests. For the Picture Concepts subtest, each point increase from Similarities predicted an increase of .32 points on the Picture Concepts test: $R^2 = .21$, $F(3, 72) = 6.48$, $p < .05$. For the Matrix Reasoning subtest, each point increase from Similarities predicted an increase of .30 points on the Matrix Reasoning test: $R^2 = .21$, $F(3, 72) = 6.48$, $p < .05$. For the Block Design subtest, Similarities was not a statistically significant predictor variable, with a 1-point increase on Similarities predicting a .00 point increase on Block Design: $R^2 = .21$, $F(3, 72) = 6.48$, $p = ns$ (See Table 5).

Correlation coefficients were computed between the Similarities subtest and each of the three PRI subtests, as well as between the VMI and each of the three PRI subtests. As explained above, the Bonferroni correction was applied to reduce the likelihood of Type I error, which suggested a $p$ value of < .0055 for significance. The results of the correlational analyses are presented in Table 6. In order to determine whether there was a statistically significant difference in the strength of the correlation relationships, correlation coefficients were compared using the methods established by Blalock and colleagues (1972), and the Bonferroni correction was again used for the six analyses such that a $p$ value of < .0083 was required for significance. The results of the comparison analyses are presented in Table 7. Results will be reported as they relate to the six specific hypotheses.
Hypothesis 1

It was predicted that the Matrix Reasoning score would correlate more highly with the Picture Concepts score than the Block Design score. Scores on the Matrix Reasoning subtest were significantly correlated with scores on the Picture Concepts subtest: \( r = 0.32 \) \( (p = 0.0043) \). The correlation between Matrix Reasoning scores and Block Design scores was also statistically significant: \( r = 0.32 \) \( (p = 0.0051) \).

The two correlation coefficients from the above comparison were then compared to one another, taking into account their shared variation. The results indicated that the correlation coefficient between Matrix Reasoning scores and Picture Concepts scores (0.32) was not significantly higher than the correlation coefficient between Matrix Reasoning scores and Block Design scores (0.32) at the 0.05 level, \( t(73) = 0.5963, ns \). Therefore, hypothesis 1 was not supported.

Hypothesis 2

It was predicted that the Similarities score would correlate more highly with the Matrix Reasoning score than the Block Design score. Similarities shared a low to moderate correlation with Matrix Reasoning: \( r = 0.36 \) \( (p = 0.0012) \). Similarities was not significantly correlated with the Block Design subtest: \( r = 0.18 \) \( (p = 0.1288, ns) \).

A comparative analysis (Blalock, 1972) taking into account shared variability was conducted with the correlation coefficients from the above analyses to evaluate whether Similarities scores were more closely related to Matrix Reasoning scores or to Block Design scores. The results indicated that the correlation coefficient between Similarities scores and Matrix Reasoning scores (0.36) was not significantly higher than the correlation coefficient
between Similarities scores and Block Design scores (.18), \( t(73) = .3383, \) \( ns \). Hypothesis 2 did not appear to be supported.

**Hypothesis 3**

It was predicted that the VMI score would correlate more highly with the Block Design score than the Matrix Reasoning score. The VMI score shared a low to moderate correlation with Matrix Reasoning: \( r = .35 (p = .0021) \). The VMI score was not significantly correlated with Block Design: \( r = .11 (p = .3530, ns) \).

A comparison analysis taking into account shared variance between the correlation coefficients (Blalock, 1972) was conducted with the correlation coefficients from the above analyses to evaluate whether VMI scores were more closely related to Matrix Reasoning scores or to Block Design scores. The results indicated that the correlation coefficient between VMI scores and Block Design scores (.11) was not significantly lower than the correlation coefficient between VMI scores and Matrix Reasoning scores (.35), \( t(73) = .2067, ns \). Hypothesis 3 did not appear to be supported.

**Hypothesis 4**

It was predicted that the Picture Concepts score would correlate more highly with the Matrix Reasoning score than with the Block Design score. As noted above, the Picture Concepts score was significantly correlated to the Matrix Reasoning score: \( r = .32 (p = .0043) \). The Picture Concepts score was not significantly correlated with the Block Design score after applying the Bonferroni correction: \( r = .25 (p = .0311, ns) \).

A comparison analysis taking into account shared variance in the correlation coefficients (Blalock, 1972) was conducted with the results of the above analyses to evaluate whether Picture Concepts scores were more closely related to Matrix Reasoning scores or to Block Design scores.
scores. The results indicated that the correlation coefficient between Picture Concepts scores and Matrix Reasoning scores (.32) was not significantly higher than the correlation coefficient between Picture Concepts scores and Block Design scores (.25), $t(73) = .0419, ns$. Hypothesis 4 did not appear to be supported.

**Hypothesis 5**

It was predicted that the Similarities score would correlate more highly with the Picture Concepts score than the Block Design score. The Similarities score was moderately correlated with the Picture Concepts score: $r = .39 (p = .0006)$. The Similarities score was not significantly correlated with the Block Design score: $r = .18 (p = .1288, ns)$.

A comparison analysis taking into account shared variance between the correlation coefficients (Blalock, 1972) was conducted with the results of the above analyses to evaluate whether Similarities scores were more closely related to Picture Concepts scores or to Block Design scores. The results indicated that the correlation coefficient between Similarities scores and Picture Concepts scores (.39) was not significantly higher than the correlation coefficient between Similarities scores and Block Design scores (.18), $t(73) = 1.0139, ns$. Hypothesis 5 did not appear to be supported.

**Hypothesis 6**

It was predicted that the VMI score would correlate more highly with the Block Design score than the Picture Concepts score. The VMI score did not appear to be significantly correlated with the Picture Concepts score for the present sample: $r = .09 (p = .4226, ns)$ nor with the Block Design score: $r = .11 (p = .3530, ns)$.

A comparison analysis taking into account shared variance between the correlation coefficients (Blalock, 1972) was conducted with the results of the above analyses to evaluate
whether VMI scores were more closely related to Picture Concepts scores or to Block Design scores. The results indicated that the correlation coefficient between VMI scores and Block Design scores (.11) was not significantly higher than the correlation coefficient between VMI scores and Picture Concepts scores (.09), \( t(73) = -1.0212, ns \). Hypothesis 6 did not appear to be supported.

**Discussion**

Intellectual assessment is utilized for a variety of purposes, including educational placement and intervention, vocational placement and rehabilitation, to aid in clinical interventions, to clarify diagnosis, and to assess the integrity of the brain after potential injury (Harrison et al., 1988). Given that intellectual assessment measures are used for a wide range of purposes in environments and in manners which can have a long-term impact on an individual’s treatment, education, and/or vocational training, it is important that the tools used to measure intelligence be capable of providing reliable and valid information about the abilities being assessed. The study of intellectual assessment methods and instruments is relevant to enhancing the tools themselves, and this may account, in part, for the decreasing duration between new editions of the gold standard instruments in intellectual assessment, the Wechsler scales.

The latest Wechsler scale for children, the WISC-IV (Wechsler, 2003) has introduced a new subtest, called the Picture Concepts subtest. It has also included a new modification of a test which was formerly introduced for adults on the WAIS (WAIS-IV; Wechsler, 2008), the Matrix Reasoning test. Although construct validity for most WISC-IV subtests has been established by studies of previous versions of the WISC, no such studies have been completed regarding the Picture Concepts and Matrix Reasoning subtests of the WISC because they remain relatively new. Therefore, the present study was conducted with the intention of providing additional
information about the construct validity of the Picture Concepts and Matrix Reasoning subtests of the WISC-IV.

The WISC-IV manual (Wechsler, 2003) directs that the Picture Concepts and Matrix Reasoning subtests be administered and interpreted as core portions of the Perceptual Reasoning Index (PRI), a composite of three subtests all purported to measure the construct of nonverbal reasoning to varying degrees. Additionally, the Matrix Reasoning subtest is generally assumed to measure visual-spatial abilities (Wechsler, 2003; Sattler, 2008). Researchers have hypothesized that the Matrix Reasoning and Picture Concepts subtests also require some degree of verbal mediation for successful completion of the task, which would suggest that an individual’s scores on these subtests would reflect, in part, the degree to which a person successfully utilized verbal mediation when arriving at the selected response to each item (Dugbartey et al., 1999; Roth, 1978; Sattler & Dumont, 2008). The present study was designed to determine whether the two target subtests appear to be more closely related to visual-spatial tasks, such as the Beery’s VMI, or to verbal tasks, such as the Similarities subtest of the WISC-IV. The Similarities subtest is one of three core subtests for the Verbal Comprehension Index of the WISC-IV. This subtest was chosen over the other two core subtests because it relies somewhat on abstract reasoning skills, and was thought to be the subtest which would most closely correlate to the other subtests, and therefore the subtest with the least likelihood of producing a Type I error. Although it was hypothesized that scores on both target subtests, Picture Concepts and Matrix Reasoning, would be more highly correlated with the Similarities subtest scores rather than the VMI scores, it was expected that the pattern would be more pronounced for the Picture Concepts scores than the Matrix Reasoning scores. This is because, although both are hypothesized to incorporate verbal
mediation, Picture Concepts directions prime participants to use verbal mediation by asking participants to justify their answers to two sample items verbally.

Unfortunately, none of the six hypotheses appeared to be supported by the present study. There were no significant relationships observed between the subtest scores other than what can be accounted for by shared variance. In other words, although all of the subtests shared variance due to them all measuring $g$ to varying degrees, no additional relationships were observed between the various subtests.

There could be a number of reasons why the above hypotheses were not supported. One possible problem would be a lack of statistical power due to a relatively small sample size ($N = 76$) and a relatively large number of statistical analyses. To determine the plausibility of this limitation, a post hoc power analysis was completed using G*Power (Faul, Erdfelder Buchner, & Lang, 2009). The present sample size was sufficient to have high power for a large effect size (0.5), adequate power for a medium effect size (0.3), and low power for a small effect size (0.1). Unfortunately, because the present study is a preliminary study, it is difficult to determine the expected effect size and whether power may have been a plausible problem. Given the shared variance among the observed scores due to all measuring sub-abilities related to $g$, it is likeliest that a small to medium effect size would be expected.

Although it appears likely that the present study had a sufficient sample size in terms of statistical power, there were differences between the sample used in the present study compared to the standardization sample of the WISC-IV and the greater population. The WISC-IV sample size was matched to the March 2000 U.S. Census in terms of racial and ethnic makeup, parental education level, and geographic region of the United States at that time. The sample was not matched in terms of age and gender for the purpose of obtaining adequate sample size within
each category. However, the age and gender distributions were reasonably matched to the census data regardless. The present sample was not matched deliberately with regard to age and race/ethnicity. Data was not collected about parental educational level, and, as described above, most of the data was collected from the Northwest region of the United States. Therefore, there are areas of difference between the present sample and the standardization sample from the WISC-IV, as well as between the present sample and the most recent census data. Specifically, the present sample was over eighty percent Caucasian, whereas the standardization sample was closer to sixty percent Caucasian (Wechsler, 2003).

In terms of score distribution, the present sample also differed from both the WISC-IV standardization sample and the theoretical population distribution. As noted previously, exclusion criteria precluded having any data from individuals with FSIQ scores below 80, in order to reduce confounding variables related to certain cognitive disorders. However, this created a necessarily skewed curve for the sample data, and resulted in an average FSIQ score for the sample of approximately 111. Subtest score averages ranged from 11.07 to 12.28 (see Table 3). Standard deviations from the present sample were closely matched to the standardization sample, and both sample populations appeared to have standard deviations somewhat below the theoretical standard deviations of 15 for Index scores and 3 for subtest scores. Although the average scores from the present sample are all within a standard deviation of the standardization sample averages and theoretical population averages, it is possible that the differences impacted the results of the study. However, these differences appear to be primarily accounted for by the exclusion of FSIQ scores below 80 rather than another source of skew.

Another area of difference between the WISC-IV data and the results of the present study is the magnitude of difference observed between the reported inter-subtest correlations from the
manual compared to the correlations obtained by the present study and reported in Table 4. Of the subtests analyzed in the present study, the inter-subtest correlations ranged from .41 to .55 in the WISC-IV technical manual (Wechsler, 2003). However, they ranged from .18 to .39 in the present sample. Although it is difficult to determine the cause of this difference, it likely relates to differences in the samples and sample sizes. For example, the present study excluded participants with FSIQ scores below 80, and collected data specifically within one region of the United States. The racial and ethnic makeup of the present sample differed from the standardization sample, and the present sample was not stratified based upon age or gender as was the standardization sample. These differences, particularly combined with differences in sample sizes, may have contributed to the differences in scores.

As another potential limit to the statistical strength of the study, a conservative approach to statistical significance was used in the present study in order to reduce the likelihood of Type I error. Using the Bonferroni approach increases the likelihood that the results will support the null hypothesis. However, it was determined that the Bonferroni approach to control for Type I error was the best option given the high number of correlations and other analyses completed. Additionally, many of the analyses which failed to reach statistical significance according to Bonferroni criteria would have also failed to reach statistical significance using another source of Type I error control, such as a .03 or .01 $p$ value. This provides additional support that the Bonferroni correction likely did not mask any true findings.

Finally, it is possible that the hypotheses are incorrect and no significant relationship of the nature proposed in this study exists between the different subtests. In other words, it is possible this study and Kain’s (2006) study failed to find significant evidence for the impact of verbal mediation on performance on the Matrix Reasoning and Picture Concepts subtests
because use of verbal mediation strategies does not highly impact performance on these subtests. In fact, there is some reason to suspect, based upon the findings from the present study, that visual-spatial processing may be more important for the Matrix Reasoning subtest in particular, than was hypothesized at the outset of the present study.

Examination of the magnitude of correlations suggests that having an increased sample size might allow a future study to find statistically significant differences among some of the comparison analyses, and this is partially supported by the results of a post hoc power analysis. However, the correlation coefficients are generally similar in magnitude, and it is not anticipated that a larger sample size would enable one to observe a clinically meaningful or relevant differences in the correlation coefficients. In other words, it is possible that verbal mediation is occurring and the present study failed to demonstrate this. However, it is unlikely that a study with a larger sample size alone would find differing effects. A follow-up study with greater statistical power would better clarify these possibilities.

The results of the Picture Concepts portion of the present study are consistent with Kain’s (2006) Picture Concepts study. Kain previously failed to find evidence for verbal mediation on the Picture Concepts subtest of the WISC-IV, and the present results are consistent with this finding. As Kain’s study of the subtest is the only current published study of construct validity for the Picture Concepts subtest to date, no other comparisons are able to be made.

The results of the Matrix Reasoning portion of the present study are somewhat consistent with published research. Previous researchers have found mixed results, with one study suggesting that verbal mediation is essential to successful completion of the Matrix Reasoning subtest (DeShon, Chan, & Weissbein, 1995), and another study suggesting that verbal mediation
is useful for some types of matrix tasks, but actually interferes with other types of matrix tasks (Welsh, 1987).

Future research in this area may benefit from consideration of the strengths and weaknesses of the present study. Strengths of the study included that the statistics were managed conservatively, and the Bonferroni correction was utilized to reduce the likelihood of making a Type I error. Additionally, the present study was inclusive of participants regardless of age, gender, race, total IQ score, etc. Therefore, results will likely be relevant for a wide variety of WISC-IV examinees. Weaknesses of the present study included a relatively small sample size, which may have impacted the statistical power. Another weakness of the study relates to the sample of participants itself. This sample was not collected in a randomized manner. Most participants were in acquaintance-level relationships with the examiners, and this may have created a possible selection bias. One possible outcome of this selection bias is the distribution of scores, which suggested that the present sample overall had significantly higher IQ scores than is generally observed in the general population. Unfortunately, it is difficult to determine to what extent this trend relates to any possible selection bias, or to what extent the trend relates to the Flynn effect, which suggests IQ scores increase over time. This is a strong possibility given that many of the participants were tested 6-8 years after the release of the WISC-IV, which is now in its eighth year. Finally, a weakness of the study relates to the high number of test administrators. Although all test administrators were thoroughly trained, and competency checks were conducted, the possibility remains that non-standardized administration may have occurred. However, this remains a potential risk within the domain of assessment as a whole, and may not have had any additional impact on the present study.
Several directions for future research are suggested by the present study. Given the developmental progression of verbal reasoning skills throughout childhood, it might be useful to determine whether verbal mediation occurs differently in children who are younger than latency age when compared to children who are latency age or older. Additionally, it would be useful to further explore the relationship between visual skills and the PRI subtests, particularly Block Design and Matrix Reasoning.

The present study was conducted to investigate the potential impact of verbal mediation on another cognitive process, namely fluid reasoning. Additionally, the study was conducted due to a lack of independent investigation of the two new WISC-IV subtests, Matrix Reasoning and Picture Concepts. Although the results of the present study were largely null findings, there are some interesting directions for future research, as noted above. Additionally, further investigation of the two new WISC-IV subtests remains important prior to accepting these subtests as valid tests of the stated abilities. Although the WISC-IV has now been in publication for over eight years, there still remains a dearth of adequate independent investigation into certain portions of the instrument. Further investigation of the test, and the two new subtests in particular, remains necessary.
References


### Table 1

**CHC Abilities and Related WISC-IV Subtests**

<table>
<thead>
<tr>
<th>CHC Ability</th>
<th>WISC-IV Subtest Related to Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluid intelligence ($G_f$)</td>
<td>Matrix Reasoning</td>
</tr>
<tr>
<td></td>
<td>Picture Concepts</td>
</tr>
<tr>
<td></td>
<td>Arithmetic</td>
</tr>
<tr>
<td>Crystallized intelligence ($G_c$)</td>
<td>Similarities</td>
</tr>
<tr>
<td></td>
<td>Word Reasoning</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
</tr>
<tr>
<td></td>
<td>Information</td>
</tr>
<tr>
<td></td>
<td>Comprehension</td>
</tr>
<tr>
<td>Short-term memory ($G_{sm}$)</td>
<td>Digit Span</td>
</tr>
<tr>
<td></td>
<td>Arithmetic</td>
</tr>
<tr>
<td></td>
<td>Letter-Number Sequencing</td>
</tr>
<tr>
<td>Visual processing ($G_v$)</td>
<td>Block Design</td>
</tr>
<tr>
<td></td>
<td>Picture Completion</td>
</tr>
<tr>
<td>Auditory processing ($G_a$)</td>
<td>None</td>
</tr>
<tr>
<td>Long-term storage and retrieval ($G_{lr}$)</td>
<td>None</td>
</tr>
<tr>
<td>Processing speed ($G_s$)</td>
<td>Symbol Search</td>
</tr>
<tr>
<td></td>
<td>Cancellation</td>
</tr>
<tr>
<td></td>
<td>Coding</td>
</tr>
</tbody>
</table>

Table 2

*Structure of WISC-IV*

<table>
<thead>
<tr>
<th>WISC-IV Index</th>
<th>Core Subtests</th>
<th>Supplemental Subtests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptual Reasoning Index</td>
<td>Block Design</td>
<td>Picture Completion</td>
</tr>
<tr>
<td></td>
<td>Matrix Reasoning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Picture Concepts</td>
<td></td>
</tr>
<tr>
<td>Verbal Comprehension Index</td>
<td>Vocabulary</td>
<td>Information</td>
</tr>
<tr>
<td></td>
<td>Similarities</td>
<td>Word Reasoning</td>
</tr>
<tr>
<td></td>
<td>Comprehension</td>
<td></td>
</tr>
<tr>
<td>Processing Speed Index</td>
<td>Coding</td>
<td>Cancellation</td>
</tr>
<tr>
<td></td>
<td>Symbol Search</td>
<td></td>
</tr>
<tr>
<td>Working Memory Index</td>
<td>Digit Span</td>
<td>Arithmetic</td>
</tr>
<tr>
<td></td>
<td>Letter-Number Sequencing</td>
<td></td>
</tr>
</tbody>
</table>
Table 3

*Subtest Score Distributions (standardization sample means and standard distributions for subtests in parentheses)*

<table>
<thead>
<tr>
<th>Subtest</th>
<th>M</th>
<th>SD</th>
<th>Min. Value</th>
<th>Max. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSIQ</td>
<td>110.95 (101)</td>
<td>11.44 (11.7)</td>
<td>81</td>
<td>132</td>
</tr>
<tr>
<td>VCI</td>
<td>111.53 (100)</td>
<td>13.66 (11.7)</td>
<td>77</td>
<td>142</td>
</tr>
<tr>
<td>Similarities</td>
<td>12.11 (10.1)</td>
<td>3.19 (2.6)</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>12.00 (10.1)</td>
<td>2.37 (2.3)</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>Comprehension</td>
<td>12.12 (10.1)</td>
<td>2.58 (2.5)</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>Picture Concepts</td>
<td>11.07 (10.1)</td>
<td>2.90 (2.7)</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>12.28 (10.2)</td>
<td>2.76 (2.5)</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>Block Design</td>
<td>11.21 (10.0)</td>
<td>2.48 (3.0)</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>Beery’s VMI</td>
<td>104.33</td>
<td>12.36</td>
<td>77</td>
<td>142</td>
</tr>
</tbody>
</table>

*Note.* All reported statistics are based upon the sample size of 76. Means and distributions for WISC-IV standardization sample obtained from the technical manual (Wechsler, 2003).
Table 4.

*Degree of Skewness and Kurtosis Observed for Key Variables.*

<table>
<thead>
<tr>
<th></th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSIQ</td>
<td>-0.24</td>
<td>-0.46</td>
</tr>
<tr>
<td>Similarities</td>
<td>-0.68</td>
<td>0.51</td>
</tr>
<tr>
<td>Picture Concepts</td>
<td>-0.49</td>
<td>0.12</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>-0.54</td>
<td>1.35</td>
</tr>
<tr>
<td>Block Design</td>
<td>-0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>Beery’s VMI</td>
<td>0.39</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Table 5

*Simple Linear Regression Model with Perceptual Reasoning Subtests as a Function of Similarities Scores*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Subtest scores as a function of Similarities score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$</td>
</tr>
<tr>
<td>Constant</td>
<td>85.84*</td>
</tr>
<tr>
<td>Picture Concepts</td>
<td>.32*</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>1.58*</td>
</tr>
<tr>
<td>Block Design</td>
<td>.00</td>
</tr>
</tbody>
</table>

$R^2$ = .21  
$F$ = 3.30

*Note. N = 76. CI = confidence interval. *$p < .05.$
Table 6

*Correlation Coefficients for Subtests Analyzed*

<table>
<thead>
<tr>
<th></th>
<th>Picture Concepts</th>
<th>Matrix Reasoning</th>
<th>Block Design</th>
<th>Similarities</th>
<th>Beery’s VMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture Concepts</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>.3237*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block Design</td>
<td>.2476</td>
<td>.3182*</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarities</td>
<td>.3854*</td>
<td>.3637*</td>
<td>.1758</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Beery’s VMI</td>
<td>.0933</td>
<td>.3471*</td>
<td>.1080</td>
<td>†</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* Correlations significant at the .0055 level are marked with an *. The *p*-value required for significance was calculated by applying the Bonferroni correction such that the usual *p*-value of .05 was divided by the number of correlations calculated for the analysis (.05/9 = .0055). † Correlation not calculated, as the correlation is not relevant to analyses.
Table 7

*T-values for Correlation Comparisons*

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Correlation Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_1$</td>
</tr>
<tr>
<td>Comparison 1</td>
<td>.32 (MR:PC)</td>
</tr>
<tr>
<td>Comparison 2</td>
<td>.36 (SI:MR)</td>
</tr>
<tr>
<td>Comparison 3</td>
<td>.35 (VMI:MR)</td>
</tr>
<tr>
<td>Comparison 4</td>
<td>.32 (PC:MR)</td>
</tr>
<tr>
<td>Comparison 5</td>
<td>.39 (SI:PC)</td>
</tr>
<tr>
<td>Comparison 6</td>
<td>.09 (VMI:PC)</td>
</tr>
</tbody>
</table>

*Note.* All statistically significant comparisons would be marked with an *. None of the comparative analyses reached statistical significance.
Figure 1. Example item from Matrix Reasoning subtest of the WISC-IV.

Figure 2. Example item from Picture Concepts subtest of the WISC-IV.