CHAPTER 12

Wechsler Nonverbal Scale of Ability (WNV)

Jack A. Naglieri, Kimberly Brunnert

INTRODUCTION

General ability (*g*) has been the underlying model for IQ tests since the early 1900s. These tests were and continue to be comprised of questions that are verbal (e.g., vocabulary or word analogies), quantitative (e.g., math word problems or math calculation), and spatial (arranging blocks to match a simple design or assembling puzzles to make a common object). The *spatial* tests have been described as *nonverbal* because it is an easier concept to understand, not because of any intention to measure *nonverbal ability*. In fact, Wechsler's view was that "the subtests are different measures of intelligence, not measures of different kinds of intelligence" (1958, p. 64). The *Technical and Interpretive Manual* for the Wechsler Nonverbal Scale of Ability (WNV; Wechsler & Naglieri, 2006a) also cites Boake (2002) as noting that "Wechsler viewed verbal and performance tests as equally valid measures of intelligence" (2006, p. 1). Further, Naglieri (2003b, 2008a) wrote that the term *nonverbal* refers to the content of the test, not a type of ability.

There is considerable experimental support for the concept of general intelligence as measured by tests such as the Wechsler and Binet (see Jensen, 1998, for a review). But the content of these tests sometimes presents a problem for assessment of culturally and linguistically diverse populations. For an individual who has not had the chance to acquire verbal and quantitative skills due to limited opportunity to learn or a disability, verbal and quantitative tests designed to measure general ability may be a good predictor of current academic performance but not a good reflection of their ability. For example, typical Native-American Navajo children living on a reservation in northern Arizona earn low scores on the Verbal scale but average scores on the Performance scale of the Wechsler (Naglieri & Yazzie, 1983) because they speak English as a second language and have had insufficient exposure to the language of a typical American child. Suzuki and Valencia (1997) argued that verbal and quantitative questions found on most traditional IQ tests interfere with accurate assessment of minority children; therefore, a nonverbal test of general ability such as the WNV (Wechsler & Naglieri, 2006b) offers a viable method for evaluating ability for these children.

The essence of a nonverbal test of general ability is that it does not contain verbal and quantitative test questions, although it may

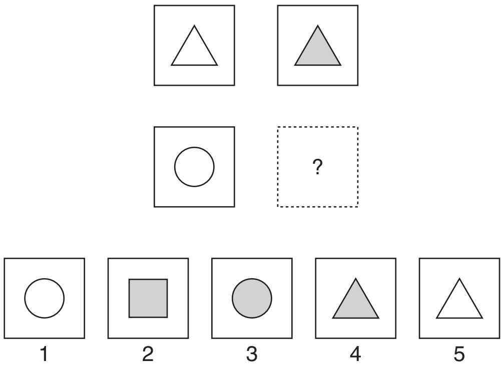


FIGURE 12.1
Simple Test Question from the WNV Matrices Subtest

involve verbal ways to solve the problem and it does require intellectual effort on the part of the examinee. For example, Figure 12.1 shows a simple nonverbal test question like those included in the Matrices subtest of the WNV. The item has shapes that vary across the horizontal and vertical dimensions that the examinee must recognize and understand to answer the analogy expressed using geometric figures (e.g., it is a figural analogy arranged in a matrix). In Figure 12.1, for example, the differences between the top and bottom rows is that the shape inside boxes changes (a triangle appears on the top row and a circle on the bottom row). The difference between the first and second column included in the top row is the addition of shading in the right column. The examinee needs to understand the interrelationships among these variables (shape and shading across the columns and rows) to arrive at the correct answer (option 3). The child may, or may not, use a verbal description (in any language) of the figures contained within the matrix or the child may simply look at the shapes and understand which option is the answer with minimal verbal analysis.

Tests that measure general ability nonverbally may have different formats, but the essential goal of these tests is the same: to measure ability nonverbally. For example, some nonverbal tests are comprised of one type of item, the progressive matrix (e.g., the Naglieri Nonverbal Ability Test—Second Edition [NNAT-2]; Naglieri, 2008b) given in a group format or individual format (Naglieri Nonverbal Ability Test—Individual Form [NNAT-1]; Naglieri, 2003b). Another method is to use several different types of nonverbal subtests as found in the WNV (as well as the Universal Nonverbal Intelligence Test [UNIT; Bracken & McCallum, 1998]; see Bracken and McCallum's chapter in this book). The slight variation in administration format and subtest composition notwithstanding, the goal is the same: to measure general ability nonverbally, and in so doing, provide a way effectively and fairly to assess a wide variety of individuals regardless of their educational or linguistic backgrounds.

DESCRIPTION OF THE WNV

Subtest Background

The WNV is comprised of a variety of subtests that are intended to measure general ability in different ways. For example, although the nonverbal subtests on the WNV are all alike in that they do not require language or arithmetic skills, they are diverse in their specific requirements. For example, some of the subtests have a strong visual-spatial requirement, others demand recall of spatial information or recall of the sequence of information, and others involve paper-and-pencil skills. This multidimensionality of task requirements distinguishes the WNV from tests that use one type of test format, such as the NNAT-2 (Naglieri, 2008b), which uses progressive matrices exclusively.

Most of the WNV subtests have appeared in previous editions of the Wechsler scales (Wechsler Preschool and Primary Scale of Intelligence—Third Edition [WPPSI-III], Wechsler Intelligence Scale for Children—Third Edition [WISC-III], Wechsler Intelligence Scale for Children—Fourth Edition

[WISC-IV], Wechsler Intelligence Scale for Children—Fourth Edition—Integrated [WISC-IV Integrated], Wechsler Adult Intelligence Scale—Fourth Edition [WAIS-III], Wechsler Memory Scale—Third Edition [WMS-III]), and have an established record of reliability and validity for the nonverbal measurement of general ability. Adaptation of the subtests was necessary to accommodate the new pictorial directions format, identify items that were most appropriate for the specific ages, and provide directions in the six languages. Each WNV subtest was included only after careful examination of both the content and form of the items vis-à-vis the goals of this particular instrument. The origins and descriptions of the WNV subtests are referenced in Table 12.1.

Each subtest is further described in the following.

Matrices

The Matrices (MA) subtest requires the examinee to discover how different geometric shapes are spatially or logically interrelated. The multiple-choice options provide potential answers that vary in the degree to which each option completes the relationships among the parts. The items are displayed using basic geometric figures such as squares, circles, and triangles using some combination of the colors black, white, yellow, blue, and green. Items were constructed using shapes and colors that would maintain interest and minimize the likelihood that impaired color

TABLE 12.1 Subtests Origin and Description

Subtest (Abbreviation)	Origin and Description
Matrices (MA)	This subtest was adapted from the NNAT–I. The examinee looks at an incomplete figural matrix and selects which of the four or five response options is the missing piece.
Coding (CD)	This subtest was adapted from the WISC-IV. The examinee follows a key that provides symbols that correspond with shapes (Coding A) or numbers (Coding B).
Object Assembly (OA)	This subtest was adapted from the WPPSI-III and the WISC–III, and has one new item. The child is presented with puzzle pieces that are placed by the examiner in a specified layout. The child completes the puzzle within a specified time limit.
Recognition (RG)	This is a new match-to-stimulus subtest. The child looks at a page with a design with geometric patterns on it for three seconds. The child then chooses which of four or five response options on the next page match the viewed stimulus.
Spatial Span (SSp)	This subtest was adapted from the WMS–III. The examinee mimics the examiner's tapping on a series of blocks either in the same order as the examiner (Spatial Span Forward) or in the reverse order (Spatial Span Backward).
Picture Arrangement (PA)	This subtest is adapted from the WAIS–III and a research version of the WISC–IV Integrated. The examinee uses a set of picture cards, which the examiner has placed on the table in a specified order, to tell a logical story within a specified time limit.

vision would influence the scores. The WNV Matrices items are composed of a variety of formats (e.g., geometric patterns, reasoning by analogy, and spatial visualization) previously used in the NNAT-I. Matrices is always administered (i.e., it is given to examinees in both age bands and is included in both the 4- and 2-subtest batteries).

Coding

The Coding (CD) subtest requires the examinee to copy symbols (e.g., a dash, two vertical lines, an open parenthesis) that are paired with simple geometric shapes or numbers according to a key provided at the top of the page. There are two forms of the Coding subtest: Form A is used in the 4-subtest battery for ages 4:0–7:11 and Form B is used in the 4-subtest battery for ages 8:0–21:11. The Coding subtest is adapted for use in the WNV from the WISC-IV by eliminating reversible shapes (e.g., left and right parentheses) for the younger age group and evenly distributing the use of each code across each row (e.g., for Coding Form B, the stimuli range from 1 to 9).

Object Assembly

The Object Assembly (OA) subtest is comprised of items that require the examinee to complete pieces of a puzzle to form a recognizable object such as a ball or a car. These items vary in the number of pieces (from 2 to 11) and the complexity with which they have been disassembled. Object Assembly is included in the 4-subtest battery of the WNV for examinees ages 4:0–7:11. The Object Assembly subtest was adapted for use in the WNV by using items from WPPSI-III (e.g., bear, apple, dog, star, calf, and tree), the WISC-III (e.g., ball), and one new item (i.e., glasses).

Recognition

The Recognition (RG) subtest was created for use in the WNV and is included in both the 4- and 2-subtest batteries for examinees ages 4:0–7:11. It requires the examinee to examine a stimulus (e.g., a square with a small circle in the center) for three seconds and then choose which option is identical to the stimulus that was just seen. The figures are colored black, white, yellow, blue, and/or green to maintain interest and minimize the likelihood that impaired color vision will influence the scores.

Spatial Span

The Spatial Span (SSp) subtest requires the examinee to touch a group of blocks arranged on an 8-by-11-inch board in a nonsystematic spatial manner in the same and reverse order of that demonstrated by the examiner. Spatial Span is included in both the 4- and 2-subtest batteries for ages 8:0–21:11. The Spatial Span subtest was adapted for use in the WNV from the WMS-III and adapted, like all the subtests, to the pictorial directions format.

Picture Arrangement

The Picture Arrangement (PA) subtest requires the examinee to arrange cartoon-like illustrations into a sequence that is logical and makes sense. Picture Arrangement is included in the 4-subtest battery for examinees ages 8:0–21:11. The Picture Arrangement subtest was adapted for use in the WNV by merging colorized items from the WAIS-III (e.g., choir, speak, and shark) and items from a research edition of the WISC-IV Integrated (e.g., duck, storm, farm, shadow, and broken).

Structure of the Test

The WNV is structured in four ways, combining subtests selected to best meet the examinee's and examiner's needs. There are 4- and 2-subtest batteries for each age band, 4:0–7:11 and 8:0–21:11. The subtests that are in each are referenced in Figure 12.2.

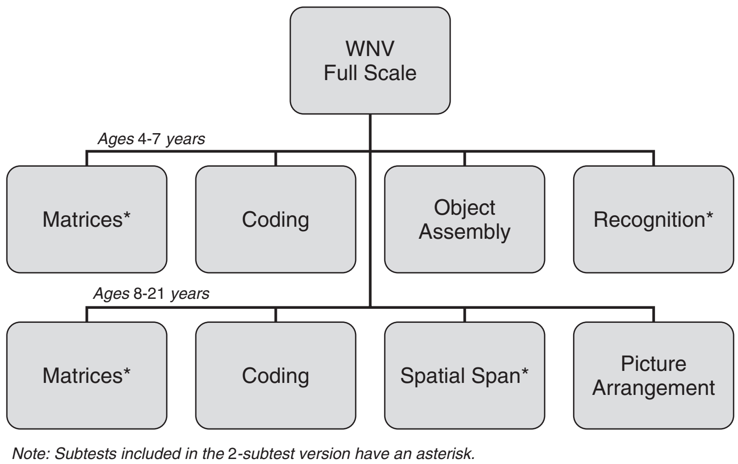


FIGURE 12.2
Structure of the WNV

The examinee's raw scores are converted to *T* scores for each subtest, which have a mean of 50 and a standard deviation of 10. Using the *T* scores for each subtest, analyses can be performed to compare the examinee's performance across the subtests and to identify strengths and weaknesses. There is a Full Scale Score that can be calculated for each battery that has a mean of 100 and a standard deviation of 15. There are separate WNV norms tables based on standardization samples collected in the United States and Canada. For Spatial Span, there are additional analyses that can be performed to examine the difference between Spatial Span Forward and Spatial Span Backward, as well as the longest span in either direction.

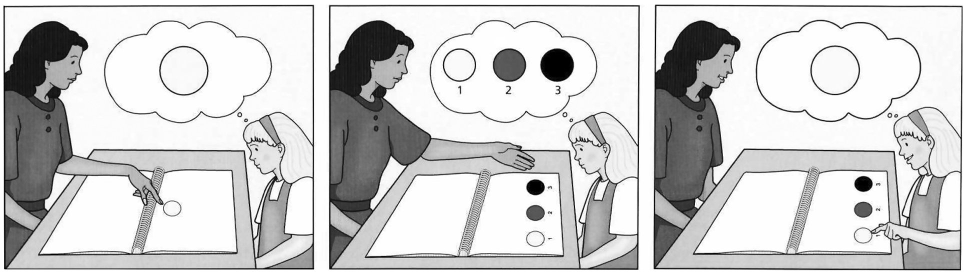
ADMINISTRATION AND SCORING

Tips on Administration

Unlike most tests of its kind, the WNV administration begins with a standardized short introduction that tells examinees that they need to look at the pictorial directions and that they can ask the examiner questions if necessary. Like the other spoken text included in the WNV, these sentences are provided in English, French, Spanish, Chinese, German, and Dutch. Actual administration procedures follow carefully scripted directions designed to ensure that the demands of the tasks are completely understood by all examinees. There are three steps to the administration directions:

Step 1 uses the standardized directions, which are always administered and should never be changed. These directions must be administered in every case and include gestures that correspond to the pictorial directions. Pictorial directions are used at Step 1 to provide a standardized method of communicating the demands of the task. These pictorial directions (see Figure 12.3) show a scene like the one the examinee is currently in. The frames of the directions show the progression of an examinee being presented with the question, then thinking about the item, and finally, choosing the correct solution.

Step 1 instructions include actions by the examiner that must be carefully followed to maximize the likelihood that the examinee understands the correspondence between the materials and the task. Gestures are used to direct the examinee's attention to specific portions



Note: The actual WNV Pictorial Directions appear in color.

FIGURE 12.3
WNV Pictorial Direction

of the pictorial directions and to the stimulus materials and sometimes to demonstrate the task itself. Sometimes simple statements are also included because they convey the importance of both time and accuracy to the examinee.

Step 2 directions are used only after the standard directions are provided. These instructions must also be followed exactly and are given only when an examinee is unclear about what he or she is being asked to do. These directions include standardized simple sentences and gestures for communicating the requirements of the task to the examinee. Verbal directions provide another way to ensure that the examinee understands the demands of the tasks and are provided in English as well as Spanish, German, French, Chinese, and Dutch. These translations are to be used only when the following two conditions are both met: (1) the WNV is being administered to an examinee who speaks one of the languages, and (2) the examiner or a professional interpreter speaks the language.

Step 3 gives an opportunity to provide help, but these directions are used only after the previous two steps have been administered. This is the only step of administration that gives the examiner flexibility. For example, the examiner may say or sign additional instructions or questions. In general, examiners are given the opportunity to communicate in whatever manner they think will best explain the demands of the subtest based on their judgment of the examinee's needs. This could include providing further explanation or demonstration of the task, restating or revising the verbal directions, or using additional words to describe the requirements of the task. At no time, however, is it permissible to teach the examinee how to solve the items. Instead, the goal of Step 3 instructions is to provide additional help to ensure that the examinee understands the demands of the task, not to show how to teach a way to solve the items. The amount of help provided and the determination about when to stop is based on professional judgment.

When using an interpreter to facilitate communication prior to and during administration, it is important that the interpreter have guidelines and training about what is and what is not permitted. This person should translate a general explanation of the testing situation for the examinee, including the introductory paragraph at the beginning of Chapter 3 in the WNV *Administration and Scoring Manual* before administration begins. It is also important that the interpreter recognize the boundaries of his or her role in administration. For example, although it is appropriate for the interpreter to translate the examiner's responses to an examinee's response to a sample item, it is not acceptable for the interpreter to make additional statements unless instructed to do so. Importantly, at no time should the interpreter communicate any information that could influence the examinee's scores. See Brunnert, Naglieri, and Hardy-Braz (2009) for more information about working with translators and especially when testing those who are deaf or hard of hearing.

Administration of the WNV subtests is designed to be simple and easy. The *Administration and Scoring Manual* includes a section prior to actual administration directions that describes the subtest, the materials needed, start, stop, and reverse rules, scoring, as well as general issues unique to each subtest. The Manual also provides considerable discussion of the physical materials, uses and applications for the two versions of the WNV, and general testing, administration, and scoring issues. In this chapter, we will highlight some of the most important points, but the reader is advised to carefully study that Administration and Scoring Manual. What follows is a subtest-by-subtest discussion of specific administration issues.

Matrices

Although the Matrices subtest is a very straightforward subtest to administer, examiners should always be aware of possible responses that may suggest concern. For example, some students who are particularly impulsive might select the option that is mostly but not completely correct because the options were written with varying degrees of accuracy. If an examinee is not looking at the options closely, one of those that is almost correct may be selected. Similarly, if an examinee takes a long time to respond, the examiner may (after about 30 seconds) prompt a response.

Coding

The examinee is allowed to correct mistakes by crossing out the incorrect symbol and writing his or her response next to it, so the pencil without an eraser needs to be used for the Coding subtest. The examiner should ensure that the examinee works from left to right and from top to bottom without skipping any items or row and by providing the appropriate instruction when needed. The examiner also provides instruction that informs the examinee to work as quickly as possible. For that reason, examinees should not be allowed to spend too much time making corrections.

Object Assembly

The examiner should always set up the puzzle pieces on the same side of the Stimulus Book as the examinee's dominant hand. Then remove the Stimulus Book before administering the Sample Item. The examiner should also ensure that the examinee works as quickly as possible. If the examinee is still completing a puzzle when the time limit expires, the examiner should place his or her hand over the puzzle to stop the examinee's progress, and record the examinee's answer. If the examinee seems upset at that point, the examiner should allow the examinee to finish but not consider any additional work for scoring purposes. It is also important to remember to begin timing after the last word of the instruction is provided. Assembling the pieces for the examinee requires a specific method fully articulated in the WNV *Administration and Scoring Manual* (Wechsler & Naglieri, 2006c). Essentially, the method requires that the puzzle pieces are put before the child in a specific format *face down*. Once all the pieces are before the child, then they are turned over in the order indicated by the number on the back of each piece.

Recognition

Examiners must be sure to expose each stimulus page for exactly 3 seconds. To do so will likely require that the page be exposed at a precise time when the stopwatch strikes an exact second and then being prepared (e.g., having your finger under the page so that you are ready) to turn the page exactly when the 3 seconds have elapsed. Do not allow examinees to turn the pages.

Spatial Span

The Spatial Span Board must be placed so that the examinee can easily reach all cubes on the board and only the examiner can see the numbers on the back of each blue block. Also, always set the Spatial Span Board on the same side of the Stimulus Book as the examinee's dominant hand. Tap the blocks at a rate of one per second and raise the hand approximately one foot above the Spatial Span Board between each tap. If the examinee does not respond after the examiner taps a sequence, the examiner can say, "It's your turn." Always administer both Spatial Span Forward and Spatial Span Backward, regardless of the examinee's performance on Spatial Span Forward and always administer both trials of an item regardless of the examinee's performance on the first trial.

Picture Arrangement

The examiner should always place the Picture Arrangement Cards on the same side of the Stimulus Book as the examinee's dominant hand and remove the Stimulus Book with Pictorial Directions before administering the Sample Item. Having the cards in the box in the order in which they are to be exposed to the examinee is an excellent way to efficiently deliver the item. When the examinee completes the item, record his or her sequence, then resequence the cards in the presentation order for the next administration. If the examinee is going very slowly, it is permissible to ensure that the examinee realizes that he or she should work as quickly as possible. If the examinee orders the cards from right to left instead of left to right, the examiner should ask, "Where does it start?"

If the examinee is in the midst of completing a story when the time limit expires, the examiner should place his or her hand over the story to stop the examinee's progress, and record the examinee's answer. If the examinee seems upset that he or she was stopped while completing the story, the examiner should allow the examinee to finish. However, the examiner should not consider any additional work by the examinee for scoring purposes.

How to Score the Test

Five of the six subtests (i.e., Matrices, Coding, Recognition, Spatial Span, and Picture Arrangement) are scored by summing the number of points earned during administration. The sixth subtest (i.e., Object Assembly) has time bonuses for some items that might be part of the raw score.

The raw scores are converted to *T* scores. The sum of *T* scores is converted to a Full Scale Score, which has a corresponding percentile rank and confidence interval.

Information can also be assessed at the subtest level for the 4-subtest battery by comparing the *T* score an examinee earned on a subtest to the mean *T* score for all four subtests administered. If the difference between the subtest *T* score and the mean *T* score is significant, then that subtest is considered a strength or weakness. Base rates are provided for each difference. Additionally, optional analyses can be performed on the Spatial Span results as described later in this chapter.

Use of Scoring and Report Writing Software

The WNV Scoring Assistant provides an automated way to obtain all standard and derived scores based on the U.S. as well as the Canadian versions of the test. The report writing feature of the software provides reports that are appropriate for clinicians as well as parents. The parent report is available in English, French, and Spanish. The software also provides links between the WNV and the WIAT-II and all the ability comparisons to achievement.

STANDARDIZATION, NORMS, AND PSYCHOMETRICS

Characteristics of the Standardization Sample

There were two samples collected for the creation of the WNV norms: one in the United

States and the other in Canada. There were also samples collected of special groups and of other tests for validity research. The U.S. sample consisted of 1,323 examinees stratified across 5 demographic variables: age (4:0–21:11), sex, race/ethnicity (Black, White, Hispanic, Asian, and Other), education level (8 years or less of school, 9–11 years of school, 12 years of school [high school degree or equivalent], 13–15 years of school [some college or associate's degree], and 16 or more years of school [college or graduate degree]), and geographic region (Northeast, North Central, South, and West). Note that education level was determined by the parent education for examinees ages 4:0–17:11 and by the examinee's own education for ages 18:0–21:11.

The Canadian sample consisted of 875 examinees stratified across five demographic variables: age (4:0–21:11), sex, race/ethnicity (Caucasians, Asians, First Nations, and Other), education level (less than a high school diploma; high school diploma or equivalent; college/vocational diploma or some university, but no degree obtained; and a university degree), and geographic region (West, Central, and East). Additionally, the Canadian sample consisted of 70% English speakers, 18% French speakers, and 12% speakers of other languages.

Reliability of the Scales

The reliability estimates for the WNV were provided by subtest and Full Scale Scores by age and over all ages. There are reliability estimates provided for the U.S. normative sample, for the Canadian normative sample, and for all the special groups that are reported in the WNV *Technical and Interpretive Manual*. The reliability estimates for the U.S. normative sample ranged from .74 to .91 for the subtests and were .91 for both Full Scale Scores across ages. The reliability estimates for the Canadian normative sample ranged from .73 to .90 for the subtests, were .90 for the Full Scale Score: 4-Subtest Battery, and .91 for the Full Scale Score: 2-Subtest Battery. The reliability estimates for the studies with examinees that were diagnosed with or classified as being gifted, having mild mental retardation, moderate mental retardation, reading and written expression learning disorders, language disorders, English language learners, Deaf, and Hard of Hearing are shown in Table 12.2.

TABLE 12.2 Reliability Estimate Ranges by Special Study

Special Group	**Reliability Estimate Range**
Gifted	.77–.97
MR Mild	.80–.93
MR Moderate	.87–.93
Reading and Written Expression Learning Disorders	.72–.88
Language Disorders	.74–.97
English Language Learners	.70–.96
Deaf	.77–.98
Hard of Hearing	.75–.97

Other reliability information is provided in the WNV *Technical and Interpretive Manual*. This other information includes the standard error of measurements (SEM), confidence intervals, and test–retest stability estimates. SEMs are provided for both the U.S. and the Canadian normative samples by subtest and Full Scale Scores by age in years, for ages 4:0–15:11, and by age bands, from ages 16:0–21:11. The confidence intervals, which are calculated with the standard error of estimate (SEE), are provided alongside the *T* Score to Full Scale Score conversion tables at both the 90% and 95% levels (see Tables A.2 and A.3 in the WNV *Administration and Scoring Manual*). Test–retest stability estimates were also provided in the WNV *Technical and Interpretive Manual*.

Use of the WNV

Interpretation Methods

Like any test, WNV test results should always be interpreted with consideration of the many factors that can influence obtained scores. Perhaps the most important are issues such as the behaviors observed during testing, and relevant educational and environmental backgrounds, and physical and emotional status, all within the context of the reason for referral. In order to obtain the greatest amount of information from the WNV, there are some important methods of interpretation that warrant discussion. Some of these methods are the same for the 4- and 2-subtest batteries, and others are unique to each version. In this chapter, the issues that apply to both batteries will be covered first and then the finer points of interpretation relevant to each version will be examined separately.

Interpretation of the Two WNV Versions

Both versions of the WNV are comprised of subtests (set at a mean of 50 and *SD* of 10) that are combined to yield a Full Scale score (set at a mean of 100 and *SD* of 15). The WNV subtest scores are set on the *T* score metric of 50 and 10 (as opposed to a traditional scaled score with a mean of 10 and *SD* of 3). This format was selected because the individual subtests had sufficient range of raw scores, allowing for the use of *T* scores, which have a greater range and precision than scaled scores. For example, the WNV subtest *T* scores range from 10 to 90, yielding 81 possible different scores, whereas a scaled score typically ranges from 3 to 20, yielding only 18 different scores. The use of the *T* score also provides greater precision on each subtest, allowing for higher reliability coefficients of the Full Scale score.

The WNV Full Scale scores are standard scores with a mean of 100 and *SD* of 15 based on either the 4- or 2-subtest batteries. This score provides a nonverbal estimate of general ability that has excellent reliability and validity. Both the Full Scale and subtest *T* scores are based on the U.S. or Canadian standardization samples and can be used to measure general ability nonverbally. It is important to recognize that the *nonverbal* label refers to the fact that the test items do not contain verbal and quantitative content and does not suggest that a specific type of ability is being assessed. Additionally, even though the WNV subtests have different demands—that is, some are spatial (e.g., Matrices or Object Assembly), others involve sequencing (Picture Arrangement and Spatial Span), require memory (e.g., Recognition and Spatial Span), or use symbol associations (Coding)—they all measure general ability. The WNV measures general ability nonverbally. General ability allows us, for example, to understand spatial as well as verbal and mathematical concepts, remember visual relationships as well as quantitative or verbal facts, and work with sequences of information of all kinds. The content of the questions may be visual or verbal, and require memory or recognition, but general ability (sometimes referred to as *g*) underlies performance on all these kinds of tasks and the WNV Full Scale is an excellent measure of *g*.

WNV 4-Subtest Battery Interpretation

Step 1: The first step in interpretation of the 4- and 2-subtest versions of the WNV is to examine the Full Scale score. This score is the most reliable and valid representation of general ability on the scale. The Full Scale score should be reported with its associated percentile score, categorical description (Average, Above Average, etc.), and confidence interval. A statement such as the following illustrates how these concepts might be included in a written document:

Sam obtained a WNV Full Scale score of 91, which is ranked at the 27th percentile and falls within the Average classification. This means that he performed as well as or better than 27% of examinees his age in the normative sample. There is a 90% chance that his true Full Scale score falls within the range 85–99.

Step 2: The second step in interpretation of the 4-subtest version of the WNV is to examine the *T* scores the examinee earned on the subtests. Analysis of this type must take into consideration the lower reliability of these scores and the increased probability that variability will reflect measurement error. Examination of the four WNV subtests should also be conducted with consideration that even though the subtests are all nonverbal measures of general ability they do have unique attributes (i.e., some involve remembering information, others spatial demands, etc.). Additionally, variability across the subtests should be expected and, therefore, statistical guidelines should be followed to ensure that any differences interpreted are beyond those that could be expected by chance. In fact, Wechsler and Naglieri (2006a) reported the cumulative percentages, mean, and *SD* of subtest scatter (e.g., highest–lowest subtest score for each individual in the U.S. standardization group) in Table B.5 in the WNV *Administration and Scoring Manual*. The mean score was 16.5 (*SD* = 7.5), indicating that practitioners can expect differences among the WNV subtests. In fact, approximately 50% of that sample had a range of subtest scores that was between 0 and 16. If *unusual* is defined at 10% of the U.S. standardization sample, then a range of 27 or more would meet that criterion. Scatter is one way to determine whether the WNV subtests vary, but when the goal is to determine whether there is significant variability and to relate the unique contributions of each subtest to other findings, a different method is recommended. That method requires that the examinee's subtest scores are compared to that examinee's mean.

The values needed for significance when comparing a WNV subtest for an examinee to that examinee's mean *T* score are provided in the WNV *Administration and Scoring Manual* (Table B.1) and in more detail by Brunnert, Naglieri, and Hardy-Braz (2009) and summarized here in Table 12.3. These values were computed utilizing Davis's (1959) formula (and Silverstein's 1982, modification of this procedure) for the difference between the average of several scores obtained by one examinee and each of his or her scores included in the average. For example, when four WNV subtest *T* scores are compared with the mean of the four *T* scores, a *z* value that takes into account the number of comparisons needs to be used. The standard errors of measurement for the WNV subtest *T* scores of the U.S. and Canadian samples were obtained from the WNV *Technical and Interpretive Manual* (Wechsler & Naglieri, 2006a, Tables 4.4 and 4.5) and were used for these calculations. (See Silverstein, 1982, for more information and Naglieri and Paolitto, 2005, for values for the WISC-IV.)

The following steps should be used to compare each of the four WNV subtest *T* scores to the child's mean subtest *T* score:

1. Calculate the mean of the four subtest *T* scores.
2. Calculate the difference between each subtest *T* score and the mean.
3. Subtract the mean from each of the subtest *T* scores (retain the sign).
4. Find the value needed for significance using the examinee's age group and the desired significance level in Table 12.3.
5. If the absolute value of the difference is equal to or greater than the value in the table, the result is statistically significant.

TABLE 12.3 Differences Required for Significance When Comparing Each WNV Subtest for the 4-Subtest Battery *T* Scores to the Examinee's Average Subtest *T* Score for the U.S. and Canadian Standardization Samples by Age Group

Country	Age	*p*-Value	MA	CD	OA	RG
United States	4:0–7:11	.10	7.4	7.9	8.7	8.6
		.05	8.3	8.8	9.7	9.6
			MA	CD	SSp	PA
	8:0–21:11	.10	6.8	8.7	7.5	9.3
		.05	7.6	9.7	8.4	10.4
			MA	CD	OA	RG
Canada	4:0–7:11	.10	8.1	8.0	9.1	8.6
		.05	9.0	8.9	10.1	9.6
			MA	CD	SSp	PA
	8:0–21:11	.10	6.8	8.7	7.4	9.6
		.05	7.6	9.7	8.3	10.7

NOTE: MA = Matrices, CD = Coding, OA = Object Assembly, RG = Recognition, SSp = Spatial Span, and PA = Picture Arrangement.

6. If the subtest difference from the mean is lower than the mean, then the difference is a weakness; if the subtest difference from the mean is greater than the mean, then the difference is a strength.

For example, if an 18-year-old from the United States obtained *T* scores of 65 on Matrices, 42 on Coding, 39 on Spatial Span, and 61 on Picture Arrangement, the mean *T* score would be 51.8. Using the values from Brunnert et al. (2009) and the .05 level of significance, Matrices would be considered a relative strength for this examinee (65 – 51.8 = 13.3, which exceeds the critical value of 7.6 for the .05 level of significance). The *T* score for Spatial Span is significantly lower than the examinee's mean *T* score (39 – 51.8 = –12.8, which exceeds the critical value of 8.4 for the .05 level of confidence), and would be considered a relative weakness. Similarly, the *T* score for Coding is significantly lower than the mean *T* score (42 – 51.8 = –9.8, which exceeds the critical value of 9.7 for the .05 level of confidence), and is also a relative weakness.

When there is significant variability in the WNV subtests, it is also important to determine whether a weakness relative to the examinee's overall mean is also sufficiently below the average range. Determining whether a child has significant variability relative to his or her own average score is a valuable way to determine strengths and weaknesses relative to the child's mean score, but Naglieri (1999) cautioned that a relative weakness could also be significantly below the normative mean. He recommended that any subtest score that is low relative to the child's means should also fall below the average range to be considered a noteworthy weakness (e.g., < 1 *SD* below the mean). In the example above, the Spatial Span *T* score of 39 would be considered a weakness from the ipsative *and* normative perspectives, but the Coding score of 42 would not. This would strengthen the

TABLE 12.4 Differences Required for Significance When Comparing Recognition or Spatial Span to Matrices on the WNV 2-Subtest Battery for U.S. and Canadian Standardization Samples by Age Group

		Age Group in Years	
		4:0–7:11	**8:0–21:11**
United States			
$p = .10$	MA vs. RG	9.9	—
	MA vs. SSp	—	5.3
$p = .05$	MA vs. RG	11.8	—
	MA vs. SSp	—	6.3
Canada			
$p = .10$	MA vs. RG	9.2	—
	MA vs. SSp	—	8.1
$p = .05$	MA vs. RG	11.0	—
	MA vs. SSp	—	9.7

level of concern about this finding and more strongly suggest that additional examination using a multidimensional measure of ability such as the Cognitive Assessment System (see Naglieri, 1999) could be appropriate. Subtest differences that are significant (in addition to the Full Scale score) should be described in a manner similar to the following:

> *His scores on the individual WNV subtest scores varied significantly, suggesting a relative strength on Matrices, a subtest that requires understanding the relationships among spatial designs. A relative weakness was found on Coding and his Spatial Span subtest* T *score was significantly below his overall subtest mean and the normative mean of 50. Both of these subtests require recall of information, and this finding suggests that further examination of immediate memory may be indicated.*

WNV 2-Subtest Battery Interpretation

The differences required for significance when each pair of WNV subtests included in the 2-Subtest battery are compared are provided in Table 12.4. These values are used, for example, to determine whether a *T* score difference of 11 points between Matrices and Recognition is significant for an examinee who is 7:3 years old. To use this table, simply subtract one subtest *T* score from the other (ignore the sign) and compare the result to the value in the table that corresponds to the desired level of significance (.10 or .05). If the obtained value is equal to or greater than the value in the table, then the result is significant. Determining how often a difference of a specific magnitude occurred in the U.S. or Canadian samples would further clarify the importance of the finding. For example, Table B.4 of the WNV *Administration and Scoring Manual* provides the base rates of subtest *T* score differences by the direction of the difference. Using that table, we find that about 11% of examinees aged 8:0–21:11 obtained *T* scores for Matrices that were 14 or more points higher than their *T* scores for Spatial Span. This information can be used to augment the interpretation of the significance of the difference between the scales.

WNV Full Scale Score Interpretation

The WNV Full Scale score is a nonverbal measure of general ability that should be reported with the corresponding confidence interval, percentile rank, and classification. The Full Scale and subtest *T* scores could be described in a manner similar to the following:

> *Gene obtained a WNV Full Scale score of 98, which falls in the Average range and is ranked at the 45th percentile. This means he did as well as or better than 45% of examinees his age in the normative sample on this nonverbal measure of general ability. There is a 90% chance that his true Full Scale score falls within the range of 91 to 105. There was significant variability between the two subtests that comprise the WNV. Gene performed significantly better on a subtest requiring reasoning with geometric designs (Matrices) than he did on a subtest requiring recall of a sequence of movements arranged in a spatial array (Spatial Span).*

This difference is also unusual, with only 4.3% of the population obtaining differences this large or larger, suggesting that further investigation is appropriate.

In-depth Interpretation of Spatial Span

The WNV Spatial Span subtest Forward and Backward scores can be interpreted separately, particularly when this test is embedded within the greater context of a comprehensive assessment. The sizes of the differences required for statistical significance by age and for the U.S. and Canadian samples are 11 and 13 for the .10 and .05 levels for the United States, and 10 and 13 for the Canadian standardization samples for the combined ages 8:0–21:11. This comparison can be accomplished using Table C.1 of the WNV *Administration and Scoring Manual*, which provides a way to convert the raw scores to *T* score equivalents for Spatial Span Forward and Spatial Span Backward. According to Table C.2 of the Manual, a difference of 9 *T* score points is needed at the .15 level (13 at the .05 level) to have a significant difference between these two scores. The frequency with which Forward and Backward score differences occurred in the normative sample are provided in Table C.3 of the test Manual, which presents the bases for the *T* score differences for the U.S. sample. The tables provide the base-rate data by the direction of the difference. For example, about 24% of examinees obtained Spatial Span Forward scores 7 or more points higher than their Spatial Span Backward scores, whereas about 25% obtained Spatial Span Backward scores 7 or more points higher than their Spatial Span Forward scores.

Comparison of the differences between Spatial Span Forward and Backward *T* scores may provide useful information, but it should be integrated within the greater context of a comprehensive assessment. For example, if a Spatial Span Forward *T* score is 14 points higher than the Spatial Span Backward *T* score for a 13-year-old U.S. examinee, the difference would be significant (only 12 points are needed at the .05 level), and the difference occurs in only 9% of the normative sample (see WNV *Administration and Scoring Manual*, Table C.3). This information would be expected to be similar to other similar test scores, such as WISC-IV Digit Span Forward vs. Digit Span Backward. Both of these tests' Backward scores should be related to the Planning Scale of the Cognitive Assessment System (see Naglieri, 1999) and may suggest that the examinee has difficulty with development and utilization of strategies for reversing the order of serial information.

IDENTIFICATION OF SPECIAL POPULATIONS

Specific Learning Disabilities

The Individuals with Disabilities Education Improvement Act (IDEIA) of 2004 has brought about widespread discussion of policies regarding the identification of children with S*pecific* L*earning* D*isabilities (SLD)*. The law no longer requires, but does not disallow, the use of an ability–achievement discrepancy for eligibility determination, but the law clearly states that a comprehensive evaluation is required and that evaluation must include tests that are not culturally or linguistically discriminatory. This need for a comprehensive evaluation has also been advocated by the National Association of School Psychologists' "Position Statement on Identification of Children with Specific Learning Disabilities" (2007). Some researchers have argued that SLD is best identified by examining a pattern of strengths and weaknesses in cognitive processing scores (e.g., Hale, Flanagan, & Naglieri, 2008; Naglieri & Kaufman, 2008). How does the WNV fit into the process of evaluating children with SLD?

The WNV provides a nonverbal measure of general ability that can be compared to current achievement test scores to help determine whether a child is demonstrating academic performance that is commensurate with ability.

Unlike other measures of general ability that contain verbal and quantitative tests, the WNV provides an evaluation of ability that is not influenced by achievement-like content (see Naglieri & Bornstein, 2003, or Naglieri, 2008a, for more discussion of the similarity of test questions on ability and achievement tests). The WNV also provides a measure of ability that can be viewed as nondiscriminatory on the basis of race, ethnicity, language, and disability (see the following sections). The WNV is not, however, designed to be a test of basic psychological processes, and other tools should be used for that purpose (e.g., the Cognitive Assessment System; Naglieri & Das, 1997).

Practitioners who wish to compare WNV scores with the Wechsler Individual Achievement Test—Second Edition (WIAT–II; Pearson, 2005) can do so using the predicted-difference and simple-difference methods. The predicted-difference method takes into account the reliabilities and the correlations between the two measures. In this method, the ability score is used to predict an achievement score, and the differences between predicted and observed achievement scores are compared. Tables B.1–B.7 in the WNV *Technical and Interpretive Manual* provide the values needed for significance when conducting this analysis for children in the United States and Tables B.8–B.14 are used for comparing scores using the WNV Canadian normative sample with the WIAT–II (see Wechsler & Naglieri, 2006a for more details).

Simple differences between the WNV 4- and 2-subtest Full Scale standard scores compared with other achievement tests are provided by Brunnert et al. (2009). The achievement tests include the Kaufman Test of Educational Achievement—Second Edition (K–TEA II; Kaufman & Kaufman, 2004), Woodcock-Johnson Tests of Achievement—Third Edition (Woodcock, McGrew, & Mather, 2001), Diagnostic Achievement Battery—Third Edition (Newcomer, 2001), and the Wide Range Achievement Test—Fourth Edition (Wilkinson & Robertson, 2006). Regardless of which method is used, the examiner can augment those findings with the base rates provided in the WNV *Technical and Interpretive Manual*'s Tables B.1–B.14 for the WIAT–II. Using both the reliability of the difference and the rate at which ability and achievement test score differences occur in the normal population will provide an effective comparison of these two measures.

Attention Deficit Hyperactivity Disorders

The attention deficit hyperactivity disorders (ADHD) are evaluated on the basis of the correspondence of behavioral characteristics described by parents with the DSM-IV-TR criteria. The essential feature of ADHD is currently described as a consistent pattern of inattention and/or hyperactivity-impulsivity found more frequently and more severely than is typical in individuals at a comparable level of development. Although the DSM-IV-TR states that those with ADHD may demonstrate variability in IQ, the diagnostic criteria are based on behavioral rather than cognitive factors even though some have argued that cognitive processing scores *should* play a role in identification (Goldstein & Naglieri, 2006). The WNV, therefore, like any test of general ability, plays a minimal role in the diagnostic process; its only role may be in helping evaluate the child's level of general ability.

Gifted

The underrepresentation of minority children in classes for the gifted has been and continues to be an important educational problem (Ford, 1998; Naglieri & Ford, 2005). In fact, Naglieri and Ford (2003) stressed that Black, Hispanic, and Native-American students are underrepresented by 50% to 70% in gifted education programs. In recent years, addressing this problem has become more focused on the types of tests used when evaluating the ability of children potentially eligible for gifted programming. Some have argued that the verbal and quantitative content of some of the ability tests used and procedures followed are inconsistent with the characteristics of culturally, ethnically, and linguistically diverse populations (Naglieri & Ford, 2005; Naglieri, Brulles, & Lansdowne, 2009). That is, because IQ has traditionally been defined within a verbal/quantitative/nonverbal framework, students with limited English-language and math skills earn lower scores on the Verbal and Quantitative scales these tests include because they do not have sufficient knowledge of the language or training in math, not because of low ability (Bracken & Naglieri, 2003; Naglieri, 2008a). If a student has not had the chance to acquire verbal and quantitative skills due to limited opportunity to learn, or a disability, verbal and quantitative tests designed to measure general ability may be a good predictor of current academic performance but an inaccurate reflection of their ability to learn especially after instruction is provided. One way to address this issue is to include tests that measure general ability nonverbally. Naglieri and Ford (2003) demonstrated the effectiveness of using a group nonverbal measure of general ability (the Naglieri Nonverbal Ability Test—Multilevel Form [NNAT-ML; Naglieri, 1997]) for increasing the identification of Hispanic and Black students. Similarly, the WNV provides an individually administered way to assess general ability nonverbally and increase the participation of minorities in gifted classes. There is evidence that children in gifted education programs earn high scores on the WNV.

The WNV was administered to gifted children who were carefully matched to control subjects included in the standardization sample on the basis of age, race/ethnicity, and education level. The differences between the means were calculated using Cohen's (1996) formula (i.e., the difference between the means of the two groups divided by the square root of the pooled variance). The study included 41 examinees, all of whom had already been identified as gifted using a standardized ability measure where they performed at 2 standard deviations above the mean or more. The students in the gifted programs performed significantly better than their matched counterparts from the normative sample with effect sizes that were large for the Full Scale Score: 4-Subtest Battery and Full Scale Score: 2-Subtest Battery. See Table 12.5 for more details.

Mentally Retarded

Naglieri and Rojahn (2001) suggested that assessment of mental retardation should take into

TABLE 12.5 WNV Means, *SD*s, and Effect Sizes for Special Populations and Matched Control Groups

	Special Population		Matched Sample			
	Mean	***SD***	**Mean**	***SD***	***n***	**Effect Size**
Moderate Mental Retardation	45.9	8.9	99.5	14.1	36	–4.5
Mild Mental Retardation	67.3	12.9	97.4	15.3	51	–2.1
Gifted	123.7	13.4	104.2	12.3	41	1.5
English Language Learners	101.7	13.4	102.1	13.4	55	0.0
Hard of Hearing	96.7	15.9	100.5	14.2	48	–0.3
Profoundly Deaf	102.5	9.0	100.8	14.3	37	0.1

consideration the appropriateness of the content of the IQ test used for the individual being evaluated. They found that minority children who were previously identified as having mental retardation earned lower Verbal than Performance IQ scores on the WISC-III, presumably because of the academic content of the verbal tests (Naglieri & Bornstein, 2003). When these same subjects were evaluated using a measure of ability that does not contain verbal and quantitative test questions (the Cognitive Assessment System; Naglieri & Das, 1997), the number of children who would still qualify as having mental retardation was reduced. They concluded, therefore, that measures of ability that do not contain verbal and quantitative questions are more desirable when assessing children for mental retardation, particularly when the children have had limited educational opportunities and have culturally and/or linguistically diverse backgrounds. The WNV is a tool that can evaluate general ability nonverbally, and thereby provide a way to assess ability without undue influence of language and knowledge for individuals suspected of having mental retardation.

There were two special studies of examinees with mental retardation performed with the standardization of the WNV, one for examinees who were diagnosed with Mild Mental Retardation and the other for examinees who were diagnosed with Moderate Mental Retardation. Examinees who were diagnosed with Mild Mental Retardation were included in the study if their cognitive ability was between 55 and 70 on a standardized ability test, if they were not institutionalized, and if they met the general criteria for inclusion in a special study (see Appendix C of the WNV *Technical and Interpretive Manual*). There were 51 examinees included in this study; 43 of these examinees were 8–21 and performed significantly worse than their matched counterparts from the normative sample with effect sizes (Cohen's *d*) large for the Full Scale Score: 4-Subtest Battery and Full Scale Score: 2-Subtest Battery. See Table 12.5 for more details.

Examinees who were diagnosed with Moderate Mental Retardation were accepted in the study if their cognitive ability was between 40 and 55 on a standardized ability test, if they were not institutionalized, and if they met the general criteria for inclusion in a special study (see Appendix C of the WNV *Technical and Interpretive Manual*). There were 31 examinees included in this study; 28 of these examinees were 8–21 and performed significantly worse than their matched counterparts from the normative sample with effect sizes (Cohen's *d*) large for the Full Scale Score: 2-Subtest Battery and Full Scale Score: 4-Subtest Battery. See Table 12.5 for more details.

Deaf and Hard of Hearing

There are numerous issues surrounding the evaluation of ability for individuals who are deaf or hard of hearing. A thorough discussion of these issues and those particularly relevant to the WNV can be found in Brunnert et al. (2009). In general, however, the assessment issues center on (1) content of the test and (2) communicating test requirements to the examinee. The former issue has been covered in the previous sections regarding evaluation of mentally retarded and gifted. The issue of communicating test requirements has also been discussed in the administration portion of this chapter. In essence, because the directions are given pictorially, and can be augmented with additional statements and/or communication using sign language, the WNV offers considerable advantages for appropriate evaluation of individuals who are deaf or hard of hearing as the research studies that follow illustrate.

Profoundly Deaf

Wechsler and Naglieri (2006a) reported a study of profoundly deaf examinees who were compared with cases from the standardization of the WNV who were matched on a number of important demographic variables. This included 37 examinees who "must not have been able to hear tones to interpret spoken language after the age of 18 months, must not lip read, must not be trained in the oral or auditory-verbal approach, and must not use cued speech (i.e., they must have routine discourse by some means of communicating other than spoken language). They must have severe to profound deafness (hearing loss measured with dB, Pure Tone Average greater than or equal to 55)" (Wechsler & Naglieri, 2006a, p. 65). Thirty-one of these examinees were 8–21 and performed minimally differently than their matched counterparts from the normative sample with effect sizes (Cohen's *d*) negligible for Full Scale Score: 4-Subtest Battery and Full Scale Score: 2-Subtest Battery. There is additional information about this sample of examinees from a survey collected with the WNV standardization (Wechsler & Naglieri, 2006a, Appendix D). See Table 12.5 for more details.

Hard of Hearing

The WNV Manual also describes a study of individuals who were hard of hearing and compared their WNV scores to a group from the standardization sample who were matched on a number of important demographic variables. This study included 48 examinees who "could have a unilateral or bilateral hearing loss or deafness, and the age of onset of their inability to hear could be any age and [they] could have cochlear implants" (WNV *Technical and Interpretive M*anual, pp. 65–66). Sixteen of these examinees were 4–7 and performed minimally differently than their matched counterparts from the normative sample with effect sizes (Cohen's *d*) that were negligible for Full Scale Score: 4-Subtest Battery, and Full Scale Score: 2-Subtest Battery. Thirty-two of these examinees were 8–20 and performed minimally differently than their matched counterparts from the normative sample with effect sizes (Cohen's *d*) that were negligible for Full Scale Score: 4-Subtest Battery, and Full Scale Score: 2-Subtest Battery. See Table 12.5 for more details; for more details about the sample, see the WNV *Technical and Interpretive Manual* (Appendix D).

English as a Second Language

It is clear that as the United States continues to become more diverse, the number of individuals whose primary language is not English will continue to increase. The largest of these groups is the Hispanic population, which is approximately 37 million or about 13% of the U.S. population, making it the largest minority group (Ramirez & de la Cruz, 2002). This population of Hispanics is dominated by individuals of Mexican origin (66.9%) who reside in the Western (44.2%) and Southern (34.8%) regions of the country. Hispanics aged 25 and older are less likely to have a high school diploma than non-Hispanic whites (57.0% and 88.7%, respectively). Importantly, 27.0% of Hispanics have less than a ninth-grade education compared with only 4.0% of non-Hispanic whites (Ramirez & de la Cruz, 2002). The large number of immigrants in this country makes clear the need for psychological tests that are appropriate for those who come from these working-class homes with parents who have limited academic and English-language skills. Nonverbal tests of general ability such as the WNV are, therefore, a particularly useful way to assess minority children because they yield smaller race and ethnic differences (which is attributed to the difference in content) while these instruments retain good correlations with achievement, and can help identify minority children for gifted programs (Bracken & McCallum, 1998; Naglieri & Ford, 2003; Naglieri & Ronning, 2000a, 2000b).

Wechsler and Naglieri (2006a) provide evidence of the utility of the WNV for individuals who are learning English. The study involved examinees who speak English as a second language who were compared to a matched sample from the WNV standardization sample. This included 55 examinees aged 8–21 years whose "native language was not English, they spoke a language other than English at home, and the examinee's parents had resided in the United States less than 6 years" (Wechsler & Naglieri, 2006a, p. 63). There were 27 Hispanics and 28 examinees who specified their primary language was Cantonese, Chinese (unspecified), Korean, Russian, Spanish, or Urdu. Additional information about this sample is available in the WNV *Technical and Interpretive Manual* (Appendix D). These examinees performed very similarly to their matched counterparts from the normative sample with negligible effect sizes (Cohen's *d*) for the Full Scale Score: 4-Subtest Battery and Full Scale Score: 2-Subtest Battery, as shown in Table 12.5.

Interventions Based on Test Results

The WNV can be used in at least two ways when a plan for instruction is being developed. At the global level, the Full Scale score provides an indication of general ability that can suggest specific instructional needs based on the characteristics of the child. For example, a child who has limited English-language skills who earns a high WNV Full Scale score as part of an assessment for possible placement in Gifted should receive these special services in an environment that recognizes his or her strengths and needs. Naglieri, Brulles, and Lansdowne (2008) provide explicit suggestions about the manner in which children who are low in academic skills but high on a nonverbal measure of general ability should be taught. Their book includes strategies that are particularly useful for diverse populations of gifted students. Instructional topics for bilingual students include critical and creative thinking, appropriate gifted educational objectives, and student-directed learning in areas such as language arts, social studies, mathematics, and science.

The specific subtest scores of the WNV, like other tests of general ability, could be used to develop hypotheses about characteristics of the student that may have implications for instruction. These hypotheses should be further evaluated by other evidence and, when possible, tested to see whether a desirable response to instruction occurs. Subtest scores on the WNV could suggest hypotheses that could be further investigated or evaluated on the basis of the child's response to intervention. For instance, a Spatial Span Forward *T* score of 54 and a Spatial Span Backward *T* score of 38 could suggest a weakness with Planning as defined by Naglieri (2008c). These two scores are both significantly different and the difference occurs in only 6.2% of the U.S. standardization sample. If this finding is corroborated by a weakness on the Planning scale of the Cognitive Assessment System (Naglieri & Das, 1997) and academic problems are found, the application of methods described by Naglieri and Pickering (2003) for teaching strategy use should be attempted. There is research that supports the value of teaching children to be strategic and the positive influence this instruction has on math and reading comprehension (see Naglieri, 2008c, for a summary).

Subtest performance on the WNV may also suggest a weakness on tests of general ability that require recall of information. The Recognition subtest requires recall of information, and if the score on that subtest is significantly lower than the child's mean and less than one *SD* from the mean (< 40), the hypothesis should be considered. Once corroborated with other test data and if accompanied by academic problems in tasks that require recall of information, teaching strategies for remembering should be implemented. This would include teaching chunking or other mnemonic methods of recall of information, for which there is considerable evidence of effectiveness (Mastropieri & Scruggs, 2006; Minskoff & Allsopp, 2003). Ways in which these methods can be communicated to teachers and parents in instructional handouts are provided by Naglieri and Pickering (2003). The WNV could, therefore, help develop possible explanations for problems associated with difficulty in the classroom.

Summary

The research studies summarized here suggest that the WNV offers an effective measure of general ability that yields expected results. Individuals identified as having mental retardation earned low scores with those with the most severe retardation earning lower scores than those with mild retardation. In contrast, children identified as being gifted earned very high scores on the nonverbal measure of general ability. Importantly, individuals who were acquiring English-language skills performed very similarly to the normative mean of 100; in fact their score was 101.7, which was very similar to the matched control group, which had a score of 102.1. Additionally, the results for individuals who are hard of hearing, as well as those for the deaf, show that these two groups also earned scores on the WNV that were very similar to the matched control groups. In summary, these data provide strong support for the use of the WNV with diverse populations.

VALIDITY

Relationships between WNV and Other Ability and Achievement Tests

The WNV is strongly related to other nonverbal measures of ability and other measures of ability that contain verbal and nonverbal scales. For example, Wechsler and Naglieri (2006a) reported that the WNV 4-subtest battery Full Scale score correlated .79 with the UNIT and .73 with the NNAT-1. Similarly, the WNV Full Scale score correlated .82 with the WPPSI-III, .82 with the WISC-IV, .84 with the WAIS-III, and .83 with the WISC-IV Spanish. These findings provide evidence that the WNV is a good test of general ability even though it does not contain verbal and quantitative subtests.

The WNV is strongly correlated with achievement (.66) as measured by the WIAT-II. The correlations between the WNV and the WIAT-II are consistent with other studies of the relationship between ability and achievement, which have yielded a correlation of about .6 (Naglieri, 1999). Previous research (e.g., Naglieri & Ronning, 2000b) with the NNAT-ML has shown a correlation of .5 to .6 for large samples of children in grades K–12. The results illustrate that the 4-subtest battery Full Scale score is effective as a predictor of academic achievement.

Demographic Differences

English Language Learners

There is good evidence that the WNV is an appropriate measure of general ability for those who have limitations in either knowledge or use of the English language. Wechsler and Naglieri (2006a) provide evidence of the utility of the WNV with examinees who have English-language limitations and hearing problems that limit their ability to acquire information. The first study included examinees whose native language was not English, where the primary language they spoke was not English, a language other than English was spoken at home, and their parents resided in the United States less than six years. When compared to a group from the normative sample matched on basic demographics, they found that the examinees learning English earned essentially the same score as the matched control of English-speaking examinees in the normative group (effect sizes for the 4- and 2-subtest batteries were .03 and .04, respectively). While these results suggested that the WNV measures general ability effectively and fairly for those with limited English-language skills when combined with the studies involving students with hearing limitations, the strength of this instrument is more clearly understood.

The findings summarized above for English language learners suggest that the WNV may be useful for addressing the underrepresentation of minority children in classes for the gifted. This has been described as one of the most important problems facing educators of gifted students (Ford, 1998; Naglieri & Ford, 2005). One solution has been to use nonverbal tests of general ability as a part of the identification procedure, particularly for children whose primary language is not English. Support for the use of a nonverbal test in this context is amply documented by Naglieri (2008a). The logic is based on the fact that traditional measures of ability include tests that require knowledge of English words and the use of language even when questions involving mathematics are used, which poses a barrier for English language learners. Suzuki and Valencia (1997) argued that these verbal and quantitative questions interfere with accurate assessment of minority children. Naglieri & Ford (2005) maintained that tests like the WNV provide an effective way to assess these individuals. Because the WNV does not penalize English language learners, it is an effective tool for assessing general ability and, therefore, should be used as part of the process to identify gifted examinees whose primary language is not English.

Deaf and Hard of Hearing

Wechsler and Naglieri (2006a) reported that the WNV Full Scale scores are also very similar for two matched samples of deaf and hearing examinees and a study with hearing-impaired examinees matched with the same demographic characteristics to hearing examinees from the normative sample. Those in the deaf sample had not ever heard spoken language; they had never heard tones after the age of 18 months, could not lip read or use cued speech, and were classified as having severe-to-profound deafness. The hard-of-hearing sample had exposure to spoken language, either through hearing or lip reading, and could have a unilateral or bilateral hearing loss or deafness. Their inability to hear could have occurred at any age and they could have cochlear implants. The results for these two groups, like the sample of examinees with English-language limitations, earned WNV Full Scale scores that were close to average and were very similar to the matched control group. The effect sizes are considered negligible and small. Taken as a whole, these studies suggest that language has a negligible effect on the WNV Full Scale score and illustrate the strength of this instrument for assessment of individuals with hearing as well as language limitations.

Gender Differences

Gender differences in ability has been a topic of considerable interest for some time, resulting in a substantial body of literature on the topic (e.g., Fennema & Sherman, 1977; Geary, 1996; Halpern, 1997). Hyde and Linn (1988) found a small mean effect size (favoring females) of .11 in verbal skills for students aged 5 through 18 years, but the differences between genders were not uniform across tasks. Halpern (1997) concluded that females outperform males on tests of verbal fluency, foreign language, fine motor skills, speech articulation, and reading and writing, but males do better on tasks that involve mental rotation, mechanical reasoning, math and science knowledge, and verbal analogies. Lynn and Irwing (2004) argued that sex differences must be viewed developmentally and with consideration of the role played by biology. Based on his research using Raven’s Progressive Matrices, Lynn (2002) argued that females are slightly better than males between the ages of 10 and 13 and that after 14 males catch up and overtake females, ending up with an advantage that reaches about 2 IQ points among adults. The differences between WNV scores by gender were recently studied by Brown (2008).

Table 12.6 provides the effect sizes for WNV Full Scale and subtest scores for 1,300 girls and boys aged 4–21 years (broken into four age groups: 4–7, 8–10, 11–14, and 15–21) who participated in the standardization sample. The results indicate that both the 4- and 2-subtest Full Scale scores showed minimal gender differences. The 4-subtest effect sizes ranged from 0.0–.16 and the 2-subtest Full Scale effect sizes ranged from .01–.10. These are negligible and inconsequential differences. Interestingly, the subtest differences were also small with the exception of the Coding subtest, which showed a female superiority of .31, .33, .17, and .48 across the four age groups. These findings indicate that the WNV Full Scale scores yield values that are very similar by gender and the differences found for Coding have little influence on the total test score.

TABLE 12.6 Male/Female Effect Sizes for the 4- and 2-Subtest Batteries Full Scales and Individual Subtests by Age in Years in the Standardization Sample

	Age (in Years)			
	4–7	**8–10**	**11–14**	**15–21**
Full Scale—4	−.16	−.06	.00	−.05
Full Scale—2	−.09	.05	.01	.10
Matrices	−.04	.08	−.06	.16
Coding	−.31	−.33	−.17	−.48
Object Assembly	−.02	—	—	—
Recognition	−.12	—	—	—
Spatial Span	—	.01	.07	.01
Picture Arrangement	—	.07	.10	.20

NOTES:
$N = 1{,}300$.
Positive effect sizes indicate male scores were higher than female scores.
Effect Size = $(X_1 - X_2) / \text{SQRT}\,[(n_1 * SD_1{}^2 + n_2 * SD_2{}^2)/(n_1 + n_2)]$.

Empirical Support for the Test Structure

The WNV offers subtest-level data as well as a Full Scale score. Empirical support for this structure was examined in three ways and reported in the Technical and Interpretive Manual. These three ways are intercorrelations, confirmatory factor analysis: communality, specificity, and error variance, and confirmatory factor analysis: model fit.

The intercorrelations for all comparisons (subtest to subtest, subtest to full-scale score, and full-scale score to full-scale score) matched the patterns reported for other Wechsler scales (e.g., WISC-IV, WPPSI-III, and WAIS-III). Additionally, the intercorrelations for the Full Scale scores were moderate to high for both age bands (4:0–7:11 and 8:0–21:11) for both the U.S. and the Canadian standardization samples.

The next measure of the strength of the test structure is comparison of the specificity and error variance as well as an examination on the loadings of each subtest on *g*. “For a subtest to provide a unique contribution to the latent variable (*g*), it is expected that the specificity will exceed the error variance” (Wechsler & Naglieri, 2006a, p. 46). In each way the specificity and error variance were reported (by age band: 4:0–7:11 and 8:0–21:11; and by smaller groupings for the older age band: 8:0–10:11, 11:0–14:11, and 15:0–21:11) the specificity exceeded the error variance, allowing the conclusion that each subtest measures something unique. An additional conclusion from the *g* loadings is that all of the subtests load on *g*.

Finally, the model fit for a single factor (*g*) model was examined. The findings indicate good fit for this model for each of the two age bands (i.e., 4:0–7:11 and 8:0–21:11). However, when the older age band was broken down, the 11:0–14:11 students did not fit as well as all other age bands, but still showed adequate fit.

Overall, the conclusion about the test structure based on empirical support is that the WNV provides useful information both at the subtest level and at a single, general factor level. Subtest scores as well as Full Scale scores are useful and important when reporting the results of the WNV.

SUMMARY

The WNV was explicitly designed to provide a nonverbal measure of general ability that would be appropriate for a wide variety of culturally and linguistically diverse populations. The selection of tests that are described as nonverbal in conjunction with Pictorial Directions and oral directions in five languages provides a unique approach to measuring general ability. The evidence provided in this chapter supports the utility of the test for fair assessment of cognitive ability of those from culturally diverse backgrounds as well as those with language differences or deficiencies as well as those who are deaf or hard of hearing. The research provided in the test manual provides a base to support the use of the instrument but additional research is needed, especially regarding the utility of the instrument within diverse clinical environments.

REFERENCES

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology*, *24*, 383–405.

Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test*. Itasca, IL: Riverside.

Bracken, B. A., & Naglieri, J. A. (2003). Assessing diverse populations with nonverbal tests of general intelligence. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 243–273). New York: Guilford.

Brown, C. (2008). Gender differences on the Wechsler Nonverbal Scale of Ability. Manuscript in preparation.

Brunnert, K., Naglieri, J. A., & Hardy-Braz, S. (2009). *Essentials of WNV Assessment*. New York: Wiley.

Cohen, B. H. (1996). *Explaining psychological statistics*. Pacific Grove, CA: Brooks & Cole.

Davis, F. B. (1959). Interpretation of differences among averages and individual test scores. *Journal of Educational Psychology*, *50*, 162–170.

Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization, and affective factors. *American Educational Research Journal*, *14*, 51–71.

Ford, D. Y. (1998). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *Journal of Special Education*, *32*, 4–14.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences*, *19*, 229–247.

Goldstein, S., & Naglieri, J. A. (2006). The role of intellectual processes in the DSM-V diagnosis of ADHD. *Journal of Attention Disorders*, *10*, 3–8.

Hale, J. N., Flanagan, D. P., & Naglieri, J. A. (2008, June). Alternative research-based methods for IDEA 2004 identification of children with specific learning disabilities. *Communique*, *36*(1), 14–15.

Halpern, D. F. (1997). Sex differences in intelligence. *American Psychologist*, *52*, 1091–1102.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*, 53–69.

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

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Test of Educational Achievement—Second Edition*. San Antonio: Pearson.

Lynn, R. (2002). Sex differences on the progressive matrices among 15–16 year olds: Some data from South Africa. *Personality and Individual Differences*, *33*, 669–673.

Lynn, R., & Irwing, P. (2004). Sex difference on progressive matrices: A meta-analysis. *Intelligence*, *32*, 481–498.

Mastropieri, M. A., & Scruggs, T. E. (2006). *Inclusive classroom: Strategies for effective instruction* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Minskoff, E., & Allsopp, D. (2003). *Academic success strategies for adolescents with learning disabilities and ADHD*. Baltimore: Brookes.

Naglieri, J. A. (1997). *Naglieri Nonverbal Ability Test—Multilevel Form*. San Antonio: Psychological Corporation.

Naglieri, J. A. (1999). *Essentials of CAS assessment*. New York: Wiley.

Naglieri, J. A. (2003a). Naglieri nonverbal ability tests: NNAT and MAT-EF. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 175–190). New York: Kluwer.

Naglieri, J. A. (2003b). *Naglieri Nonverbal Ability Test—Individual Form*. San Antonio: Psychological Corporation.

Naglieri, J. A. (2008a). Traditional IQ: 100 years of misconception and its relationship to minority representation in gifted programs. In J. VanTassel-Baska (Ed.), *Critical issues in equity and excellence in gifted education series alternative assessment of gifted learners* (pp. 67–88). Waco, TX: Prufrock Press.

Naglieri, J. A. (2008b). *Naglieri Nonverbal Ability Test—Second Edition*. San Antonio: Pearson.

Naglieri, J. A. (2008c). Best practices in linking cognitive assessment of students with learning disabilities to interventions. In A. Thomas &

J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 679–696). Bethesda: NASP.

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

Naglieri, J. A., Brulles, D., & Lansdowne, K. (2009). *Helping gifted children learn*. San Antonio, TX: Pearson.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. Austin, TX: ProEd.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing under-representation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly*, *47*, 15–160.

Naglieri, J. A., & Ford, D. Y. (2005). Increasing minority children's participation in gifted classes using the NNAT: A response to Lohman. *Gifted Child Quarterly*, *49*, 29–36.

Naglieri, J. A., & Kaufman, A. S. (2008). IDEIA 2004 and specific learning disabilities: What role does intelligence play? In E. Grigorenko (Ed.), *Educating individuals with disabilities: IDEIA 2004 and beyond* (pp. 165–195). New York: Springer.

Naglieri, J. A., & Paolitto, A. W. (2005). Ipsative comparisons of WISC-IV Index scores. *Applied Neuropsychology*, *12*, 208–211.

Naglieri, J. A., & Pickering, E. B. (2003). *Helping children learn: Intervention handouts for use at school and home*. Baltimore: Brookes.

Naglieri, J. A., & Rojahn, J. (2001). Evaluation of African-American and White children in special education programs for children with mental retardation using the WISC-III and Cognitive Assessment System. *American Journal of Mental Retardation*, *106*, 359–367.

Naglieri, J. A., & Ronning, M. E. (2000a). Comparison of White, African-American, Hispanic, and Asian children on the Naglieri Nonverbal Ability Test. *Psychological Assessment*, *12*, 328–334.

Naglieri, J. A., & Ronning, M. E. (2000b). The relationships between general ability using the NNAT and SAT Reading Achievement. *Journal of Psychoeducational Assessment*, *18*, 230–239.

Naglieri, J. A., & Yazzie, C. (1983). Comparison of the WISC-R and PPVT -R with Navajo children. *Journal of Clinical Psychology*, *39*, 598–600.

Newcomer, P. (2001). *Diagnostic Achievement Battery—Third Edition*. Austin, TX: ProEd.

Pearson. (2005). *Wechsler Individual Achievement Test—Second Edition*. San Antonio, TX: Pearson.

Ramirez, R. R., & de la Cruz, G. (2002). *The Hispanic population in the United States: March 2002*. Current Population Reports, 20-545, U.S. Census Bureau, Washington, DC.

Silverstein, A. B. (1982). Two- and four-subtest short forms of the Wechsler Adult Intelligence Scale—Revised. *Journal of Consulting and Clinical Psychology*, *50*, 415–418.

Suzuki, L.A., & Valencia, R. R. (1997). Race-ethnicity and measured intelligence. *American Psychologist*, *52*, 1103–1114.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.

Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: Pearson.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale—Fourth Edition*, San Antonio, TX: Pearson.

Wechsler, D. (1997b). *Wechsler Memory Scale—Third Edition*. San Antonio, TX: Pearson

Wechsler, D. (2002). *Wechsler preschool and primary scale of intelligence—Third Edition*. San Antonio, TX: Pearson.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*, San Antonio, TX: Pearson.

Wechsler, D., Kaplan, E., Fein, D., Kramer, J., Morris, R., Delis, D.C., & Maerlender, A. (2004). *Wechsler Intelligence Scale for Children—Fourth Edition—Integrated*. San Antonio, TX: Pearson.

Wechsler, D. & Naglieri, J. A. (2006a). *Wechsler Nonverbal Scale of Ability Technical and Interpretive Manual*. San Antonio, TX: Pearson.

Wechsler, D. & Naglieri, J. A. (2006b). *Wechsler Nonverbal Scale of Ability*. San Antonio, TX: Pearson.

Wechsler, D. & Naglieri, J. A. (2006c). *Wechsler Nonverbal Scale of Ability Administration and Scoring Manual*. San Antonio, TX: Pearson.

Wilkinson, G. S., & Robertson, G. J. (2006). *WRAT4 Wide Range Achievement Test professional manual* (4th ed.). Lutz, FL: Psychological Assessment Resources.

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