Unveiling the WISC-IV

Donna Rury Smith, Ed.D.
The Psychological Corporation

Agenda

- A broad look at the changes and why we made them.
- A closer look at the new and familiar subtests
- Research design of the new WISC IV
- Psychometric properties of the test
- Interpretive considerations

Role of Intelligence Tests

The task of assessing a child’s intelligence necessarily involves more than simply obtaining his or her scores. As Wechsler (1975) noted: What we measure with tests is not what tests measure— not information, not spatial perception, not reasoning ability. These are only a means to an end. What intelligence tests measure is something much more important; the capacity of an individual to understand the world about him and his resourcefulness to cope with its challenges.
Why Revise?

- New research on cognitive abilities, enormous literature on the WISC – III
- Demographic Shifts
  - e.g. Hispanic population changes (from 11% to 15%)
  - Regions (growth of the West/South at expense of NE)
- Flynn Effect

Revision Goals - I

- Strengthen Four-Factor Model
- Improve assessment of:
  - Fluid Reasoning
  - Working Memory
  - Processing Speed
- Enhance clinical utility, and provide strong evidence of clinical validity.

Revision Goals - II

- Remove time-bonuses where possible.
- Put speed where it belongs.
- Improve Psychometric Properties
  - Subtest floors and ceilings
  - Remove potentially biased items
- Link to measures of achievement, memory, adaptive behavior, emotional intelligence, giftedness and a cognitive process instrument.
Changes

- Updated theoretical foundations
- Enhanced clinical utility
- Increased developmental appropriateness
- Improved psychometric properties
- Increased user friendliness

Structure of WISC IV

Subtests

- Core Subtests (10 of them)
  Are administered when composite scores are desired
- Supplemental Subtests (5 of them)
  Extend the range of cognitive skills sampled and provide additional clinical information as they enable the clinician to complete additional discrepancy analyses. They can also be used as substitutes for core subtests.
Full Scale IQ

- Stronger contributions of working memory and processing speed
- 30% each VCI and PRI
- 20% each PS and WM

Composite Scores

Perceptual Reasoning Index (PRI)
- Shift in emphasis from organization to reasoning
- Emphasis on fluid reasoning in the perceptual domain
- Subtests are highly g-loaded

Composite Scores

Working Memory Index
- Essential component of fluid reasoning and other higher order skills
- Closely related to achievement and learning

See Fry & Hale, 1996; Perlow, Juttuso, & Moore, 1997; Swanson, 1996
Composite Scores

- Processing Speed Index
  - Dynamically related to mental capacity, reading performance & development, and reasoning by conservation of resources (e.g., efficiency)

  See Fry & Hale, 1996; Kail, 2000; Kail & Hall, 1994; Kail & Salthouse, 1994; Berninger, 2001

Process Scores

- In addition to the subtest and composite scores, several additional process scores which provide more detailed information about a child's performance, are available.
- No additional administration procedures are required to derive these scores.
- Process scores can NEVER be substituted for core or supplemental subtest scores in the calculation of composite scores.

User Friendliness

- Testing time reduced
- Administration procedures simplified
- Use of supplemental subtests for a core subtest based on clinical need and appropriateness
- Manual reorganization
- Record Form reorganization
Content Changes

- Deleted 3 subtests
  - Picture Arrangement
  - Object Assembly
  - Mazes
- Retained, but revised, 10 subtests
- Added 5 new subtests

New Subtests

- Picture Concepts
- Letter-Number Sequencing
- Matrix Reasoning
- Cancellation
- Word Reasoning

Picture Concepts

For each item, the child is presented with 2 or 3 rows of pictures and chooses one picture from each row to form a group with a common characteristic.

- Measure of fluid reasoning and abstract categorical reasoning (without verbal response).
- Items progress from relatively concrete to more abstract.
- 28 items
Letter-Number Sequencing

The child is read a sequence of numbers and letters and recalls the numbers in ascending order and the letters in alphabetical order.

- Measure of working memory
- Adapted from the WAIS-III (but new items)
- Involves sequencing, mental manipulation, attention, short-term auditory memory, visuospatial imaging, and processing speed
- Gets credit with either order

Matrix Reasoning

The child looks at an incomplete matrix and selects the missing portion from 5 response options. (35 items)

- Measure of fluid reasoning and perceptual organization
- Reliable estimate of general intellectual ability
- 4 types of items to assess skills
  - Continuous and discrete pattern completion
  - Classification
  - Analogical reasoning
  - Serial reasoning

Cancellation

The child scans both a random and structured arrangement of pictures and marks target pictures within a specified time limit.

- Measure of processing speed and visual selective attention
- 2 forms (Random & Structured)
- Forms share identical target locations
- Targets are animals
- Foils are common non-animal objects
Word Reasoning

The child is asked to identify the common concept being described in a series of clues.

- Measure of verbal comprehension, analogical and general reasoning ability, verbal abstraction, domain knowledge, the ability to integrate and synthesize different types of information, and the ability to generate alternative concepts.
- Designed to measure fluid reasoning with verbal material.
- 24 items

Updates to Existing Subtests

Verbal Comprehension Subtests

Similarities
Vocabulary
Comprehension
Information
Word Reasoning
### Similarities

- 23 items, 11 of which are new
- Scoring criteria for all items were revised
- Sample item has been revised to require a credible response from the child before beginning the subtest rather than the corrective feedback provided on the WISC III.
- The number of 1 point items was reduced from 5 to 2.
- Age-defined start points were added.

In what way are a ____ and a ____ alike?

### Vocabulary

- Picture naming items in the stimulus book provide more floor.
- Vocabulary words are now displayed in Stimulus Book, in addition to being read aloud.
- 36 items, including 4 new picture items and 32 verbal items (5 new, 27 retained)
- Scoring criteria for all verbal items were revised.

### Comprehension

- 21 items, 11 of them are new
- Scoring criteria for all items were revised
Information

- Is now a supplemental subtest
- 33 items, 11 of them are new
- Scoring criteria for some retained items have been modified

Perceptual Reasoning Subtests

- Block Design
- Picture Concepts
- Matrix Reasoning
- Picture Completion

Block Design

- 14 items, 11 were retained from WISC III, 3 were added to extend the ceiling
- Instructions have been shortened
- Discontinue rule has been increased from 2 to 3 consecutive scores of 0
- Time bonus scores restricted to the last 6 items (total raw scores for many young children will not include any time bonus points)
- Process scores available for untimed performance (BDN)
Picture Completion

- Now a supplemental subtest
- All new artwork
- Items are designed to focus less on attention to minor visual details and more on reasoning
- Scoring criteria were revised and include distinctions between those verbal responses that require pointing and those that should be awarded credit without pointing
- 38 items, including 13 new items

Working Memory Subtests

Digit Span
  Letter-Number Sequencing
  Arithmetic

Digit Span

- Is now a core subtest and requires administration of both Digits Forward and Digits Backward
- 8 items in Digit Span Forward (2 trials each) and 8 items in Digit Span Backward (2 trials each)
- 15 trials were retained in Digit Span Forward, and 12 were retained in Digit Span Backward
- A 2-digit sample item was added to DSB
- Process scores are available to compare DSF and DSB and LDSF and LDSB
Arithmetic

- Reduced math knowledge requirements while increasing the working memory demands.
- 34 items, 11 items involve the same math calculation as items on WISC III, but verbatim instructions have been revised.
- Time bonuses are gone.
- Task more like ‘real-world’ working memory demands.

Processing Speed Subtests

Coding
Symbol Search
Cancellation

Coding

- Verbatim instructions for both forms A and B have been shortened to be more age appropriate.
- Additional process scores will be on the WISC IV PI, but can “test the limits” now.
Symbol Search

- Search A and B were retained from WISC III, but 15 items were added to Symbol Search B to improve the ceiling for older children.
- Verbatim instructions for both forms have been shortened.

Standardization of the WISC-IV

WISC Standardization

- Ages 6 – 16
- Standardization Sample n = 2200
- 200 children at each age year
- National sample stratified by sex, parent education level, ethnicity and region.
- Extensive validity studies with 15 clinical groups.
- Links to several major tests
- See Tables 3.2-3.5 in Technical Manual for percentage descriptions of Standardization sample
### Evidence of Reliability

#### Internal Consistency

- Improved coefficients for retained subtests. Significant improvement overall of subtests from WISC III.
- Average reliability coefficients (calculated with Fisher’s z transformation) across age groups:
  - Verbal Comprehension: .94
  - Perceptual Reasoning: .92
  - Working Memory: .92
  - Processing Speed: .88 (based on split-half)
  - Full Scale: .97

### Special Group Reliabilities

- Based on sample of 661 children in 16 clinical groups. Table 4.2 in Tech Manual provides coefficients by special group.
- The majority of the subtest reliability coefficients across special groups are similar or higher to those coefficients reported for the normative sample.
- Average subtest reliability coefficients range from a low of .82 on DSF to a high of .93 on LN and on MR.
Standard Errors of Measurement

- At subtest level overall average SEMs range (across ages) from low of .97 on LN to a high of 1.38 on Cancellation.
- At index level overall average SEMs range (across ages) from low of 2.68 on Full Scale to a high of 5.21 on Processing Speed.
- As a result of relatively high reliability estimates of the composite scores, the confidence intervals tend to be small and can be found in Tables A.2-A.6 in the Administration Manual.

Test-Retest Stability

- Based on a sample of 243 children, with 18-27 from each of the 11 age groups.
- Time interval ranged from 13 to 63 days, with a mean interval of 32 days.
- Table 4.4 in Tech Manual indicate that scores possess adequate stability across time for all 5 age groups. The average corrected stability coefficient is highest for Vocabulary (.92), good (in the .80s) for BD, S, DS, CD, LN, MR, C, SS, PC, I, WR; and adequate for all others (.70s).

Test-retest Gains

- less pronounced on the Verbal Comprehension and Working Memory subtests
- score differences for combined age groups (primarily due to practice effects)
  - VCI 2.1 points
  - PRI 5.2 points
  - WMI 2.6 points
  - PSI 7.1 points
  - FSIQ 5.6 points
Interscorer Agreement

- Based on sample of 60 cases scored independently by 4 scorers with no previous WISC IV scoring experience
- Interscorer reliabilities were:
  - .98 for Similarities
  - .98 for Vocabulary
  - .95 for Comprehension
  - .96 for Information
  - .97 for Word Reasoning

Score Differences

- A statistically significant difference between scores refers to the likelihood that obtaining such a difference by chance is very low if the true difference between the scores is 0. The level of significance reflects the level of confidence you can have that the difference is a true difference.
- The difference between scores required for significance is computed from the standard error of measurement of the difference.
- Table B.1 in the Administration manual lists the differences between index scores required for statistical significance at the .15 and .05 levels of significance by age groups and by overall standardization sample.

Frequency of Index Score Differences

- The prevalence or frequency of an observed score difference in the general population is called the base rate.
- Sometimes the difference between an individual’s index scores is statistically significant but is not infrequent. The statistical significance of differences between scores and the rarity of the difference are two different issues and have two different implications for test interpretation.
- Base rates are reported in Table B.2 of the Administration manual by overall standardization sample and by ability level. It is also reported by the direction of the difference.
Subtest Differences

- The interpretation of the difference between a single subtest score and the child's own mean score is an intraindividual comparison. Strengths and weaknesses identified in this way are relative to this child's own ability level.
- Table B.5 of the Administration manual reports the minimum differences between a single scaled score and the average scales scores of various groups of subtests required for statistical significance at the .15 and .05 levels.
- Remember a difference can be statistically significant but not especially unusual. Table B.5 also provides data on the estimated base rates of the general population.

Subtest Differences

- A difference between scaled scores on a pair of subtests (pairwise comparison) can be determined to be statistically significant (use Table B.3) while base rates for differences between scores are reported in Table B.4.
- Intersubtest scatter reflects the variability of a child's scaled scores across the subtests. Such variability is often considered as diagnostically significant. Table B.6 provides cumulative percentages of intersubtest scatter within various WISC IV composite scales.
- Process score differences may be of particular interest for clinical reasons. Table B.9 presents minimum scaled score differences required for statistical significance. Base rate data is reported in Table B.10.

Evidence of Validity
### Relationship to Other Measures

Mean Scores for WISC IV and WISC III  
N=244 children, aged 6—16, mean interval 28 days

<table>
<thead>
<tr>
<th>WISC IV</th>
<th>WISC III</th>
<th>Corrected r₁₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCI/VIQ</td>
<td>103.0</td>
<td>105.4</td>
</tr>
<tr>
<td>PRI/PIQ</td>
<td>103.9</td>
<td>107.3</td>
</tr>
<tr>
<td>WMI/FDI</td>
<td>101.5</td>
<td>103.0</td>
</tr>
<tr>
<td>PSI/PSI</td>
<td>102.7</td>
<td>108.2</td>
</tr>
<tr>
<td>FSIQ/FSIQ</td>
<td>104.5</td>
<td>107.0</td>
</tr>
<tr>
<td>VCI/VCI</td>
<td>102.9</td>
<td>106.0</td>
</tr>
<tr>
<td>PRI/POI</td>
<td>103.9</td>
<td>106.9</td>
</tr>
</tbody>
</table>

### Relationship to Other Measures

Mean Scores for WISC IV and WPPSI III  
N=182 children, aged 6—7, mean interval 22 days

<table>
<thead>
<tr>
<th>WISC IV</th>
<th>WPPSI III</th>
<th>Corrected r₁₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCI/VIQ</td>
<td>100.5</td>
<td>100.2</td>
</tr>
<tr>
<td>PRI/PIQ</td>
<td>102.6</td>
<td>102.0</td>
</tr>
<tr>
<td>WMI</td>
<td>99.6</td>
<td></td>
</tr>
<tr>
<td>PSI/PSQ</td>
<td>103.1</td>
<td>104.4</td>
</tr>
<tr>
<td>FSIQ/FSIQ</td>
<td>102.7</td>
<td>102.5</td>
</tr>
<tr>
<td>GLC</td>
<td>97.1</td>
<td></td>
</tr>
</tbody>
</table>

### Relationship to Other Measures

Mean Scores for WISC IV and WAIS III  
N=198 children, aged 16, mean interval 22 days

<table>
<thead>
<tr>
<th>WISC IV</th>
<th>WAIS III</th>
<th>Corrected r₁₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCI/VIQ</td>
<td>97.3</td>
<td>100.2</td>
</tr>
<tr>
<td>PRI/PIQ</td>
<td>98.9</td>
<td>102.3</td>
</tr>
<tr>
<td>WMI/WM</td>
<td>98.7</td>
<td>97.7</td>
</tr>
<tr>
<td>PSI/PSI</td>
<td>99.5</td>
<td>102.8</td>
</tr>
<tr>
<td>FSIQ/FSIQ</td>
<td>98.5</td>
<td>101.6</td>
</tr>
<tr>
<td>VCI/VCI</td>
<td>97.3</td>
<td>100.8</td>
</tr>
<tr>
<td>PRI/POI</td>
<td>98.9</td>
<td>103.8</td>
</tr>
</tbody>
</table>
Relationship to Other Measures

Mean Scores for WISC IV and WASI
N=260 children, aged 6—16, mean interval 29 days

<table>
<thead>
<tr>
<th></th>
<th>WISC IV</th>
<th>WASI</th>
<th>Corrected r₁₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCI/VIQ</td>
<td>98.9</td>
<td>102.6</td>
<td>.85</td>
</tr>
<tr>
<td>PRI/PIQ</td>
<td>101.4</td>
<td>104.0</td>
<td>.78</td>
</tr>
<tr>
<td>WMI</td>
<td>100.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSI/PSI</td>
<td>98.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSIQ/FSIQ-4</td>
<td>100.2</td>
<td>103.6</td>
<td>.86</td>
</tr>
<tr>
<td>FSIQ/FSIQ-2</td>
<td>100.2</td>
<td>102.0</td>
<td>.83</td>
</tr>
</tbody>
</table>

Relationship to Other Measures

Mean Scores for WISC IV and WIAT II
N=550 children, aged 6—16, mean interval 12 days

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VCI and Total Achievement</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>PRI and Total Achievement</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>WMI and Total Achievement</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>PSI and Total Achievement</td>
<td>.58</td>
<td></td>
</tr>
<tr>
<td>FSIQ and Total Achievement</td>
<td>.87</td>
<td></td>
</tr>
</tbody>
</table>

Relationship to Other Measures

See Table 5.15 in Technical manual
- Picture Concepts correlations range from low (.30 for Oral Expression) to moderate (.43 for Listening Comp)
- Letter-Number Sequencing correlations were moderate across WIAT II subtests ranging from .39 (Oral Expression) to .60 (Math Reasoning)
- Matrix Reasoning correlations ranged from .42 (Oral Expression) to .59 (Math Reasoning)
- Cancellation correlated minimally with all WIAT II subtests
- Word Reasoning correlations ranged from .62 (Listening Comprehension) to .42 (Oral Expression)
Clinical Studies

Each clinical case is matched demographically to normal cases based on age, sex, ethnicity, parent education level, and geographic region.

The mean of all normal cases that match each clinical case is used.

Clinical Matched Controls

- Mental Retardation – Mild
- Mental Retardation – Moderate
- Learning Disabled – Reading
- Learning Disabled – Reading, Writing
- Learning Disabled – Reading, Writing, Math
- Learning Disabled – Math
- Receptive Language Disorder
- Receptive/Expressive Language Disorder
### Clinical Studies

<table>
<thead>
<tr>
<th>ADD/LD Combined</th>
<th>WISC-IV</th>
<th>WISC-IV PI</th>
<th>BROWN ADD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor Impaired</td>
<td>WISC-IV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hearing Impaired</td>
<td>WISC-IV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism/Aspergers</td>
<td>WISC-IV</td>
<td>WISC-IV PI</td>
<td></td>
</tr>
<tr>
<td>TBI – Open</td>
<td>WISC-IV</td>
<td>WISC-IV PI</td>
<td>NEPSY</td>
</tr>
<tr>
<td>TBI – Closed</td>
<td>WISC-IV</td>
<td>WISC-IV PI</td>
<td>NEPSY</td>
</tr>
<tr>
<td>Gifted</td>
<td>WISC-IV</td>
<td>D-KEFS</td>
<td>EQ</td>
</tr>
</tbody>
</table>

### Mental Retardation Study-1

<table>
<thead>
<tr>
<th>Mild MR</th>
<th>Matched Control</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>VCI</td>
<td>67.1</td>
<td>9.1</td>
</tr>
<tr>
<td>PRI</td>
<td>65.5</td>
<td>10.3</td>
</tr>
<tr>
<td>WMI</td>
<td>66.8</td>
<td>11.1</td>
</tr>
<tr>
<td>PSI</td>
<td>73.0</td>
<td>11.6</td>
</tr>
<tr>
<td>FSIQ</td>
<td>60.5</td>
<td>9.2</td>
</tr>
</tbody>
</table>

### Mental Retardation Study-2

<table>
<thead>
<tr>
<th>Moderate MR</th>
<th>Matched Control</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>VCI</td>
<td>52.3</td>
<td>7.5</td>
</tr>
<tr>
<td>PRI</td>
<td>52.5</td>
<td>9.2</td>
</tr>
<tr>
<td>WMI</td>
<td>57.0</td>
<td>9.5</td>
</tr>
<tr>
<td>PSI</td>
<td>58.2</td>
<td>11.0</td>
</tr>
<tr>
<td>FSIQ</td>
<td>46.4</td>
<td>8.5</td>
</tr>
</tbody>
</table>
Closer look at MR

1. Must consider both intellectual functioning and adaptive behavior.
2. The prevalence of large and unusual discrepancies between verbal and nonverbal composite scores has been shown to decrease with decreasing levels of ability (Slate, 1995; Spruill, 1996, 1998; Wechsler, 1997). Thus, there appears to be less variability in performance at both the composite (index) and the subtest levels for children with MR than for children in general population.
3. Children with MR also have relative strengths and weaknesses. Study with WISC III suggested that children with Mild MR may perform slightly better on measures of processing speed than on measures of verbal or perceptual-organization abilities (Wechsler, 1991).

<table>
<thead>
<tr>
<th>LD - Reading</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RD</td>
<td>Match</td>
<td>Diff</td>
<td>Signif</td>
</tr>
<tr>
<td>VCI</td>
<td>91.9</td>
<td>100.9</td>
<td>9.0</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>PRI</td>
<td>94.4</td>
<td>99.3</td>
<td>4.9</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>WMI</td>
<td>87</td>
<td>99.8</td>
<td>12.8</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>PSI</td>
<td>92.5</td>
<td>98.6</td>
<td>6.16</td>
<td>p&lt;.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LD – Reading &amp; Writing</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RWD</td>
<td>Match</td>
<td>Diff</td>
<td>Signif</td>
</tr>
<tr>
<td>VCI</td>
<td>94.8</td>
<td>101.3</td>
<td>6.5</td>
<td>p&lt;.000</td>
</tr>
<tr>
<td>PRI</td>
<td>98.0</td>
<td>101.0</td>
<td>3.1</td>
<td>p&lt;.000</td>
</tr>
<tr>
<td>WMI</td>
<td>90.2</td>
<td>100</td>
<td>9.8</td>
<td>p&lt;.000</td>
</tr>
<tr>
<td>PSI</td>
<td>90.6</td>
<td>102</td>
<td>11.4</td>
<td>p&lt;.000</td>
</tr>
</tbody>
</table>
## LD – Math

<table>
<thead>
<tr>
<th></th>
<th>MD</th>
<th>Match</th>
<th>Diff</th>
<th>Signif</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCI</td>
<td>93.2</td>
<td>99.3</td>
<td>6.1</td>
<td>p&lt;.000</td>
<td>.61</td>
</tr>
<tr>
<td>PRI</td>
<td>87.7</td>
<td>97.2</td>
<td>9.6</td>
<td>p&lt;.000</td>
<td>.80</td>
</tr>
<tr>
<td>WMI</td>
<td>92.9</td>
<td>99.7</td>
<td>6.7</td>
<td>p&lt;.000</td>
<td>.56</td>
</tr>
<tr>
<td>PSI</td>
<td>90.6</td>
<td>95.6</td>
<td>5</td>
<td>p&lt;.000</td>
<td>.36</td>
</tr>
</tbody>
</table>

## Related Assessments in Development

- WISC IV Integrated provides a number of standardized subtests and scores
- Allows testing-of-limits supported by normative data
- Multiple-choice version of Verbal subtests and Block Design
- Spatial Span—Spatial Working Memory
- Elithorn Mazes—Executive Function
- Frequency of error types

---

Copyright 2003 Harcourt Assessment, Inc.  
All rights reserved.
WISC-IV Spanish Edition

- Normed entirely in the U.S. on various Spanish-speaking populations
- Items adapted from WISC-IV
- Where adaptation is impossible, parallel items have been created.
- Available DECEMBER 2004