“Intelligent” intelligence testing with the WJ IV cognitive battery

Kevin S. McGrew, PhD

Institute for Applied Psychometrics & University of Minnesota

© Institute for Applied Psychometrics, Dr. Kevin S. McGrew, 12-28-15

Links to more complete sets of handouts and/or PPT slides will be provided the day of the workshop
• Institute for Applied Psychometrics (IAP)-Director

• University of Minnesota - Visiting Professor (Educ. Psych.)

• Interactive Metronome - Director of Research and Science (External Consultant) *

• Darhma Berkmana Foundation (YDB; Indonesia) – Intelligence expert for development of first Indonesian CHC-based intelligence battery for children

* Conflict of interest disclosure: Financial relationship and interest in IM; Coauthor of WJ III and WJ IV (royalty interest)
“Intelligent” intelligence testing with the WJ IV cognitive battery

- General introduction and workshop logistics
- Intelligence testing in the “big picture” context
- Brief overview of Kaufman’s “intelligent” testing approach
- Foundational empirical knowledge—“romancing the stones” (tests)
  - The WJ IV/CHC Periodic Table of Cognitive Test Elements
  - WJ IV variation and comparison procedures - brief
  - Test/cluster score difference (% base rate) rules-of-thumb

© Institute for Applied Psychometrics, Dr. Kevin S. McGrew, 12-28-15
“Intelligent” intelligence testing with the WJ IV cognitive battery

Will be covered concurrently with aid of case study

- WJ IV published & new supplemental/clinical test groupings
- WJ IV assessment trees
  - Within-CHC domain assessment trees ("drilling down")
  - Academic domain referral-focused assessment trees
- Miscellaneous topics and tidbits
- Conclusions and Q/A

© Institute for Applied Psychometrics, Dr. Kevin S. McGrew, 12-28-15
Waves Of Intelligence Test Interpretation
(Kamphaus et al., 1997)

• Wave 1 - Quantification of a General Level ($g$)

• Wave 2 - Clinical Profile Analysis

• Wave 3 - Psychometric Profile Analysis

• Wave 4 - Applying Theory to Intelligence Test Interpretation

Picture Completion

Empirical Analysis

Reliability:
- Split-half: .77
- Test-retest: .81

$g$ loading: .60 (Fair)

Subtest specificity vs. error variance (Ample)
- Most related to:
  - Block Design: .52
  - Object Assembly: .49

- Least related to:
  - Coding: .18
  - Digit Span: .25

Proportion of Variance Attributed to:
- Factor 1. Verbal Comprehension: 14%
- Factor 2. Perceptual Organization: 28%
- Factor 3. Freedom from Distractibility: 1%
- Factor 4. Processing Speed: 1%
- Abilities other than the 4 factors: 33%
- Error: 23%

Proportion of Variance When 2 Factors Are Rotated:
- Factor 1. General Verbal Ability: 15%
- Factor 2. General Nonverbal Ability: 25%

Abilities Shared with Other Subtests (Unique abilities are asterisked)

INPUT
Visual perception of meaningful stimuli (people—things)

INTEGRATION/STORAGE

Perceptual Organization (Factor Analysis: 4-Factor and 2-Factor)
Gv—Broad Visual Intelligence (Horn)
Holistic (right-brain) processing
Cognition and Evaluation of figural stimuli (Guilford)
Spatial (Bannatyne)
Simultaneous processing
Distinguishing essential from nonessential details
Visual organization without essential motor activity
*Visual recognition and identification (long-term memory)

OUTPUT
Simple motor (pointing) or vocal

Subject to Influence of:
- Ability to respond when uncertain
- Alertness to the environment
- Cognitive style (field dependence—field independence)
- Concentration
- Negativism ("Nothing's missing")
- Working under time pressure
Wave 4:
Applying Theory to Test Interpretation
(and research & development)

PASS → CAS/CAS II

Sim/Suc → KABC/KABC-II

Gf-Gc → KAIT
CHC → KABC-II

PASS → CAS/CAS II

CHC (Gf-Gc) → WJ-R/III/IV

CHC (Gf-Gc) → SB5
FIGURE 4.1   A Conceptual Model of the Variables Considered in Test Interpretation (WISC-III Picture Completion Example)

Note: There are additional narrow abilities in the domains of Gc and Gv that are not included in this figure; the rectangle represents the total score variance of the WISC-III Picture Completion test; the italicized terms represent the test characteristic information that is presented for the Wechsler Scales in Table 4.2 and in the Wechsler Scale summary pages.
### Characteristics of the Wechsler Intelligence Scales

**Battery:** WISC-III  
**Test:** Block Design  
**Age Range:** 6 to 16 years

Description of test: The examinee is required to replicate a set of modeled or printed two-dimensional geometric patterns using two-color cubes. This is a timed test.

#### Basic Psychometric Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>g loading</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>g loading a</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Specificity</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Specificity a</td>
<td>15</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Item gradients</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>20-24</td>
<td>25</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>70</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>84</td>
<td>85+</td>
<td></td>
</tr>
</tbody>
</table>

#### Test Floor

<table>
<thead>
<tr>
<th>Ages 6:0 to 6:3</th>
</tr>
</thead>
</table>

#### Test Ceiling

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
</table>

---

Sample ITDR summary page from McGrew & Flanagan (1997)
**Gf-Gc CLASSIFICATIONS** (Broad: stratum II / Narrow: stratum I)

**Visual Processing (Gv):** The ability to generate, perceive, analyze, synthesize, manipulate, transform, and think with visual patterns and stimuli (Empirical: strong).

- **Spatial Relations (SR):** Ability to rapidly perceive and manipulate visual patterns or to maintain orientation with respect to objects in space (probable).
- **Visualization (Vz):** Ability to mentally manipulate objects or visual patterns and to "see" how they would appear under altered conditions (possible).

---

**OTHER VARIABLES THAT MAY INFLUENCE TEST PERFORMANCE**

<table>
<thead>
<tr>
<th>Background and Environmental</th>
<th>Individual and Situational</th>
<th>Degree of Linguistic Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>Reflectivity/impulsivity</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>Field dependence/independence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Flexibility/inflexibility</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Planning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ability to perform under time pressure</td>
<td></td>
</tr>
</tbody>
</table>
The intent of the intelligent testing model was and remains to “bring together empirical data, psychometrics, clinical acumen, psychological theory, and careful reasoning to build an assessment of an individual leading to the derivation of an intervention to improve the life circumstances of the subject” (Reynolds, 2007, p. 1133) – in Fletcher-Janzen (2009)
The gold standard for clinical-psychometric test interpretation

Incorporates both quantitative and qualitative analysis

The first system of test interpretation that followed scientific principles and at the same time overtly sought to reduce inappropriate use of obtained test scores

Demands a very high standard of clinical expertise

The central point of intelligent testing is that the clinician’s judgement regarding the patient is the central point
“Tests do not think for themselves, nor do they directly communicate with patients. Like a stethoscope, a blood pressure gauge, or an MRI scan, a psychological test is a dumb tool, and the worth of the tool cannot be separated from the sophistication of the clinician who draws inferences from it and then communicates with patients and professionals”

If you give a monkey a Stradivarius violin and you get bad music......

You don’t blame the violin !!!!
We are the instrument !!!!
“Intelligent” intelligence testing and interpretation requires ... knowing thy instruments

An “intelligent” clinician understands and “romances the stones (tests)” which have different and multiple facets

External criterion relations
Neuropsych. interpretation

CHC ability factor classifications
Level/type of cognitive processing (Type 1 v Type 2)
Cognitive operations

Info. Proc. stimulus & response characteristics (e.g., BIS)

Error variance (reliability)
Uniqueness (specificity)
g-loading
Degree of cognitive complexity
Degree of cultural loading
Degree of linguistic demand
Ability domain cohesion
Exec. Functions/Attentional control
...most disciplines have a common set of terms and definitions (i.e., a standard nomenclature) that facilitates communication among professionals and guards against misinterpretations. In chemistry, this standard nomenclature is reflected in the Table of Periodic Elements. Carroll (1993a) has provided an analogous table for intelligence.....

(Flanagan & McGrew, 1998)
A Good Taxonomy
The Cattell-Horn-Carroll (CHC) model is the contemporary consensus taxonomy of human cognitive abilities.
Richard Snow (1993):
“John Carroll has done a magnificent thing. He has reviewed and reanalyzed the world’s literature on individual differences in cognitive abilities... no one else could have done it... it defines the taxonomy of cognitive differential psychology for many years to come.”

John Horn (1998):
A “tour de force summary and integration” that is the “definitive foundation for current theory” (p. 58). Horn compared Carroll’s summary to “Mendelyev’s first presentation of a periodic table of elements in chemistry” (p. 58).
The Cattell-Horn-Carroll (CHC) Periodic Table of Human Abilities

### Domain-Independent Capacities*
- **Gf**: I, RG, RQ
- **Gwm**: WM, MS, AC
- **Glr**: MA, MM, M6
- **Gs**: P, N, R9
- **Gt**: R1, R2, R4, R7, IT
- **Gps**: R3, PT, MT

### Acquired Knowledge Systems*
- **Gc**: LD, VL, K0, LS, CM, MY
- **Gkn**: KL, K1, K2, A5, MK, KF, LP, BC
- **Grw**: V, RD, RC, RS, WA, SG, EU, WS
- **Gq**: KM, A3

### Sensory-Motor Domain-Specific Abilities*
- **Gv**: Vz, SR, MV, CS, SS, CF, IM, PI, LE, IL, PN
- **Ga**: PC, US, UM, U8, UR, U1, U9, UP, UL
- **Gh**: (No well supported cognitive Gh & Gk narrow abilities have been identified)
- **Go**: OM

### Sensory

- **Gk**: PI, P2, P3, P4, P6, P7, P8, A1

### Motor

- **Gp**: PI, P2, P3, P4, P6, P7, P8, A1

---

* Broad ability | Glr-Learning efficiency | Narrow ability | Glr-Retrieval fluency

* Three major domain categories based on Schneider & McGrew (2012)
All information based on analysis of WJ IV norm data from ages 6 thru 19

Relative degree of cognitive complexity:
- **High**
- Medium (M/M)
- Low

<table>
<thead>
<tr>
<th>Test name abbreviation</th>
<th>NmSeries</th>
<th>(RQ)</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>g-loading*</td>
<td>.93</td>
<td>.63</td>
<td>H .80</td>
</tr>
<tr>
<td>Specificity*</td>
<td>.63</td>
<td>.73</td>
<td>.64</td>
</tr>
<tr>
<td>Reliability*</td>
<td>.63</td>
<td>.73</td>
<td>.64</td>
</tr>
<tr>
<td>CHC broad factor loading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test name abbreviation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIS content/stimulus characteristic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHC narrow ability code(s)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median correlations with R, M, W clusters</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[* high, med, low – as per Kaufman (1979) & McGrew & Flanagan (1998)]
Exploratory MDS of WJ IV norm subjects ages 6-19
WJ IV test 2D MDS (Ages 6 to 19; n = 4,082)

www.iapsych.com/mimap.pdf
Relative degree of cognitive complexity:
- High
- Medium (M/M)
- Low

All information based on analysis of WJ IV norm data from ages 6 thru 19

Test name abbreviation: NmSeries (RQ)

BIS content/stimulus characteristic: H

CHC narrow ability code(s): #

CHC broad factor loading: .93 .63 H .80

Reliability*: .63 .73 .64

Specificity*: .93 .63 H .80

Median correlations with R, M, W clusters: .63 .73 .64

g-loading*: .93 .63 H .80

Reliability*: .63 .73 .64

Specificity*: .93 .63 H .80

Median correlations with R, M, W clusters: .63 .73 .64

[High, med, low – as per Kaufman (1979) & McGrew & Flanagan (1998)]
<table>
<thead>
<tr>
<th>Reliability</th>
<th>Description</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>The degree to which a test score is free from errors of measurement. Test</td>
<td>Important for making accurate educational and/or diagnostic decisions.</td>
</tr>
<tr>
<td></td>
<td>score precision.</td>
<td>Test scores are sufficiently reliable and can be used to make diagnostic decisions.</td>
</tr>
<tr>
<td>Medium</td>
<td>Coefficients from .80 to .89 inclusive.</td>
<td>Test scores are moderately reliable and can be used to make screening decisions or can be combined with other tests to form a composite with “high” reliability.</td>
</tr>
<tr>
<td>Low</td>
<td>Coefficients below .80.</td>
<td>Test scores are not sufficiently reliable and cannot be used to make important screening or diagnostic decisions. Need to be combined with other tests to form a composite with “medium” or “high” reliability.</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td>The portion of a test’s score variance that is reliable and unique to the test.</td>
<td>A test with high specificity may be interpreted as measuring an ability distinct within a battery of tests.</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>High</td>
<td>A test’s unique reliable variance is equal to or above 25% of the total test variance and it exceeds error variance (1-reliability).</td>
<td>A test with medium specificity should be interpreted cautiously as measuring an ability distinct within a battery of tests.</td>
</tr>
<tr>
<td>Medium</td>
<td>When a test meets only one of the criteria for High.</td>
<td>A test with low specificity should not be interpreted as representing a unique ability but may prove useful in interpretation when it is considered as part of a composite or cluster of other similar tests.</td>
</tr>
<tr>
<td>Low</td>
<td>When a test does not meet either of the criteria for High.</td>
<td></td>
</tr>
</tbody>
</table>
Reliability

<table>
<thead>
<tr>
<th>NmSeries (RQ)</th>
<th>H</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>.93</td>
<td>.63</td>
<td>.80</td>
</tr>
</tbody>
</table>

Specificity

<table>
<thead>
<tr>
<th>H</th>
<th>(RQ)</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>.63</td>
<td>.73</td>
<td>.64</td>
</tr>
</tbody>
</table>

Reliable variance

Specificity – reliable unique variance
Reliability

NmSeries (RQ) #

Specificity

.93 .63 H .80

H .63 .73 .64

.Gv.

.66

.83 .60 M .70

Visual (Vz)

.H .38 .49 .38

PicRec (MV)

H .71 .47 L .50

.H .36 .25 .36
<table>
<thead>
<tr>
<th><strong>g-loading</strong></th>
<th>Each test’s loading on the first unrotated factor or component in principal factor or component analysis with all other tests from a specific intelligence battery.</th>
<th>Important indicator of the degree to which a test of an individual battery measures general intelligence. Aids in determining the extent to which a test score can be expected to vary from other scores within a profile.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
<td>General factor or g loading of .70 or higher.</td>
<td>Tests with high g loadings are not expected to vary greatly from the mean of the profile and are considered good indicators of general intelligence.</td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td>A loading of .51 to .69.</td>
<td>Tests with medium g loadings may vary from the mean of the profile as tests with this classification are considered fair indicators of general intelligence.</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>A loading of .50 or lower.</td>
<td>Tests with low g loadings can be expected to vary from the mean of the profile as tests with this classification are considered poor indicators of general intelligence.</td>
</tr>
</tbody>
</table>
IQ test battery subtest \textit{g-loadings} or saturation

General intelligence \( (g) \)

High \( g \)

Intelligence test battery test \( g \) (general intelligence) \textit{loadings (weights)}

Derived from factor analysis

Think of a \textbf{general intelligence pole} that is saturated with \textit{more g-ness} (like magnetism) at the top and \textit{less g-ness} at the bottom

Factor analysis orders the tests on the pole based on their \textit{saturation of g-ness}

(1a) Spearman’s general Factor model

© Institute for Applied Psychometrics; Kevin McGrew 1-18-15
<table>
<thead>
<tr>
<th></th>
<th>$g$</th>
<th>$h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic</td>
<td>0.81</td>
<td>0.66</td>
</tr>
<tr>
<td><strong>Phonological Processing</strong></td>
<td><strong>0.81</strong></td>
<td><strong>0.66</strong></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.80</td>
<td>0.64</td>
</tr>
<tr>
<td>Oral Vocabulary</td>
<td>0.80</td>
<td>0.64</td>
</tr>
<tr>
<td>Information</td>
<td>0.79</td>
<td>0.62</td>
</tr>
<tr>
<td>Concept Formation</td>
<td>0.78</td>
<td>0.61</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>0.75</td>
<td>0.56</td>
</tr>
<tr>
<td>Similarities</td>
<td>0.74</td>
<td>0.55</td>
</tr>
<tr>
<td><strong>Verbal Attention</strong></td>
<td><strong>0.73</strong></td>
<td><strong>0.53</strong></td>
</tr>
<tr>
<td>Block Design</td>
<td>0.71</td>
<td>0.50</td>
</tr>
<tr>
<td>General Information</td>
<td>0.71</td>
<td>0.50</td>
</tr>
<tr>
<td>Number Series</td>
<td>0.70</td>
<td>0.49</td>
</tr>
<tr>
<td>Numbers Reversed</td>
<td>0.69</td>
<td>0.48</td>
</tr>
<tr>
<td>Comprehension</td>
<td>0.69</td>
<td>0.48</td>
</tr>
<tr>
<td>Letter-Number Sequencing</td>
<td>0.68</td>
<td>0.46</td>
</tr>
<tr>
<td>Digit Span</td>
<td>0.65</td>
<td>0.42</td>
</tr>
<tr>
<td><strong>Object-Number Sequencing</strong></td>
<td><strong>0.64</strong></td>
<td><strong>0.41</strong></td>
</tr>
<tr>
<td>Picture Concepts</td>
<td>0.63</td>
<td>0.40</td>
</tr>
</tbody>
</table>

First (unrotated) principal component for WJ IV COG & WISC-IV tests ($n=173$)

<table>
<thead>
<tr>
<th></th>
<th>$g$</th>
<th>$h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual-Auditory Learning</td>
<td>0.62</td>
<td>0.38</td>
</tr>
<tr>
<td>Nonword Repetition</td>
<td>0.62</td>
<td>0.38</td>
</tr>
<tr>
<td>Symbol Search</td>
<td>0.62</td>
<td>0.38</td>
</tr>
<tr>
<td>Analysis-Synthesis</td>
<td>0.61</td>
<td>0.37</td>
</tr>
<tr>
<td>Number-Pattern Matching</td>
<td>0.59</td>
<td>0.35</td>
</tr>
<tr>
<td>Story Recall</td>
<td>0.58</td>
<td>0.34</td>
</tr>
<tr>
<td>Pair Cancellation</td>
<td>0.58</td>
<td>0.34</td>
</tr>
<tr>
<td>Visualization</td>
<td>0.55</td>
<td>0.30</td>
</tr>
<tr>
<td>Picture Recognition</td>
<td>0.49</td>
<td>0.24</td>
</tr>
<tr>
<td>Letter-Pattern Matching</td>
<td>0.48</td>
<td>0.23</td>
</tr>
<tr>
<td>Coding</td>
<td>0.47</td>
<td>0.22</td>
</tr>
<tr>
<td>Cancellation</td>
<td>0.42</td>
<td>0.18</td>
</tr>
</tbody>
</table>
IQ test battery subtest *g*-loadings or saturation

General intelligence (g) ➔

(1a) Spearman’s general Factor model

© Institute for Applied Psychometrics; Kevin McGrew 1-18-15
CHC narrow factor classification

\[0.93 \quad 0.63 \quad H \quad 0.80\]

NmSeries

H (RQ) #

\[0.63 \quad 0.73 \quad 0.64\]
Ages 9-13 in technical manual

CHC broad factor loading (average)

NmSeries

(NM) #

.93 .63  .80

.63 .73 .64
What is (relational) cognitive complexity?
Cognitive Variables in Series Completion

Thomas G. Holzman
Georgia State University
James W. Pellegrino
University of California, Santa Barbara
Robert Glaser
University of Pittsburgh

The cognitive determinants of number series completion performance were studied by presenting a systematic set of problems to college adults and to average- and high-IQ elementary-school children. In each group a combination of process and content-knowledge variables accounted for more than 70% of the variance in solution difficulty. Solution difficulty was most affected by the amount of information to be coordinated in working memory while assembling and applying the pattern description rule for the sequence. Adults could effectively coordinate more information than children, but IQ levels did not differ on this component ability. Skill in dealing with unusual, hierarchical relations and arithmetic computation also affected performance and discriminated between age and IQ levels. Comparisons with results from other types of rule-induction tasks suggested some general abilities of importance to rule induction.

Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology

Graeme S. Halford
Department of Psychology, University of Queensland, Brisbane, Queensland 4072, Australia
Graeme.Halford@uq.edu.au

William H. Wilson
Department of Computer Science & Engineering, University of New South Wales, Sydney, New South Wales, Australia
wil@csse.unsw.edu.au
www.csse.unsw.edu.au

Steven Phillips
Information Science Division, Biosciences Laboratory, Tokyo, 279, Japan
steve@sdei.jpl.nasa.gov

How Many Variables Can Humans Process?

Graeme S. Halford, Rosemary Baker, Julie E. McClelland, and John D. Bahn

1University of Queensland, Brisbane, Australia, and 2Griffith University, Brisbane, Australia
Two major classes of cognitive complexity theories (Bertling, 2012)

• **Empirical**: Post-hoc purely data-driven theories (e.g., Marshalek, Lohman & Snow, 1983)
  - $g$-loadings
  - Proximity to center of MDS spatial maps

• **Cognitive theories**: Working memory theories and the constraints placed on reasoning (e.g., $G_f$). Increasing processing demands results in an increase (demand/load) on cognitive resources
Two types of cognitive theories regarding cognitive complexity

Load placed on working memory by a task

- Focus is on the sheer number of elements or element relations in a task

Relational Complexity theory (RC): The relational complexity of a task (e.g., Birney et al., 2006); Halford, 1993; Halford et al., 1988; Just & Carpenter, 1992)

- Focus is on the complexity of the interrelated elements (pieces of information) that need to be processed in parallel
Another example is provided by Sweller (1993), who analysed the following problem: *Suppose five days after the day before yesterday is Friday. What day of the week is tomorrow?* Despite our expertise in reasoning about days of the week, this problem is frustratingly difficult. The reason is that, especially in the first sentence, numerous elements are related to each other and cannot be considered meaningfully in isolation. These relations have to be at least partially processed in order to segment the statement into subproblems that can be processed serially. The processing load is felt most keenly when we try to plan this procedure.
The **processing load** (demand on resources) imposed by interacting components of a task can be captured with the concept or **relational complexity** (Bertling, 2012; Birney et al., 2006; Halford et al., 1998)

- The key is the **number of interacting variables** (elements; arguments) that must be represented in **parallel** to implement the process
- Conceptually RC is similar to the number of factors in an experimental design
Processing complexity

• May depend on executive functions.
• The strategy used by a person may differ across people or within the same person at different times.
• The optimal strategy may not be the one that is best theoretically or as generated by an artificial intelligence (AI) algorithm.
• Individuals operate in ways that are different from theoretically optimal algorithms.
WJ IV cognitive complexity design approach based on work of Lohman & Larkin (2011)

- Increase the information processing demands of the tests within a specific narrow CHC domain.
- Design tests that place greater demands on:
  - Cognitive information processing (cognitive load)
  - Greater allocation of key cognitive resources (working memory or attentional control)
  - The involvement of more cognitive control or executive functions

© Institute for Applied Psychometrics; Kevin McGrew 1-18-15
WJ IV test 2D MDS (Ages 6 to 19; n = 4,082)
© Institute for Applied Psychometrics (IAP)
Dr. Kevin McGrew 01-20-16

Tests closest to the center are considered more cognitively complex.
Cognitive complexity interpretation aid

Gwm tests

- **H** = High
- **M** = Moderate
- **L** = Low

© Institute for Applied Psychometrics; Kevin McGrew 01-06-16
Cognitive complexity by CHC domain interpretation aid

- **H** = High
- **M** = Moderate
- **L** = Low Mod.
- **L** = Low

© Institute for Applied Psychometrics; Kevin McGrew 01-06-16
Distances between tests intended to reflect relative hypothesized differences (not quantified) along two axis.

Linguistic/language dimension classifications based on inspection of correlations with other WJ IV tests of Gc and Ga and Flanagan & Ortiz (2015) linguistic demand classifications.

Hypothesized degree of linguistic or language-domain demand.

- **Sentence Repetition**
- **Nonword Repetition**
- **Memory for Words**
- **Verbal Attention**
- **Object-Number Sequencing**
- **Understanding Directions**
- **Numbers Reversed**

**Gwm A** - tasks that make greater use of the articulatory rehearsal maintenance mechanism (Camos, 2015)

Verbal/linguistic working memory

**Gwm B** - tasks that make greater use of the attentional refreshing maintenance mechanism (Camos, 2015)

Attentional control working memory

Hypothesized degree of central-executive control network (cog. load; attentional control; degree of relational cognitive complexity)
Cognitive complexity

NmSeries (RQ) #

.93 .63 H .80

.63 .73 .64
WJ IV and WISC-IV 2D MDS solutions (n=173)

- **H** = High
- **M** = Moderate
- **L** = Low
Correlation = .93; but correspondence diverges as test become higher in g and cognitive complexity.
Median test correlations with R, M, & W clusters for ages 6-19 (not reported in technical manual)

For example, correlation of Number Series with Broad Reading cluster
External criterion relations validity

NmSeries

\[
\begin{align*}
\text{Median correlation} & \\
\text{with WJ IV reading,} & \\
\text{math, and written} & \\
\text{language} & \\
\text{achievement clusters} & \\
\text{(ages 6-19)} & \\
.93 & .63 & H & .80 \\
\hline
H & (RQ) & \# & \\
.63 & .73 & .64 &
\end{align*}
\]
Faceted models of intelligence

Auditory has recently been proposed to be added

Figure 18.3 The Berlin Model of Intelligence Structure (BIS)
Fluid and Crystallized Intelligence and the Berlin Model of Intelligence Structure (BIS)

André Beauducel and Martin Kersting

Working-memory capacity explains reasoning ability—and a little bit more

Heinz-Martin Süß*, Klaus Oberauer, Werner W. Wittmann, Oliver Wilhelm, Ralf Schulze
Cognitive operations and content dimensions

Fig. 1. The BIS. Four functional abilities are cross-classified with three content abilities. General intelligence represents the integral of all operational and content-related abilities, respectively.
Auditory intelligence: Theoretical considerations and empirical findings

Kristin Conzelmann a,b,⁎, Heinz-Martin Süss a

a Department of Psychology I, Psychological Methods, Assessment, and Evaluation Research, Otto-von-Guericke University of Magdeburg, Magdeburg, Germany
b Department of Aviation and Space Psychology, DLR German Aerospace Center, Hamburg, Germany

ABSTRACT

In the last few years, auditory intellectual abilities have received increased attention in different fields of research. However, most intelligence models have yet to include an auditory factor. This paper aimed to replicate the general auditory factor and examined whether and how the hierarchical and faceted Berlin Intelligence Structure model (BIS; Jäger, 1982) should be extended by adding an auditory dimension. Two studies included 126 students (Study 1) and a heterogeneous group of 175 adults (Study 2). Participants took a broad auditory intelligence test and the BIS test and provided a self-report of musical training. Confirmatory factor analyses revealed two separate auditory content factors: nonverbal and speech. Auditory nonverbal ability was clearly distinct from academic intelligence, whereas auditory speech ability could be completely subsumed under verbal reasoning. We suggest that auditory ability – as represented by auditory nonverbal tests – needs to be added to the BIS as an additional content dimension.

© 2015 Elsevier Inc. All rights reserved.
There has been an explosion of research on auditory abilities since Carroll’s (1993) seminal work (Schneider & McGrew, 2012). A wide-ranging collection of Ga characteristics have been related to disorders of reading, speech, and language. For example, Ga abilities are now recognized as playing a pivotal scaffolding role in the development of language and general cognitive abilities (Conway, Pisoni, & Kronenberger, 2009).
<table>
<thead>
<tr>
<th>Cognitive Operations</th>
<th>Verbal</th>
<th>Auditory</th>
<th>Figural</th>
<th>Numeric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasoning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creativity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The BIS intelligence framework (2016)
Another perspective:
BIS framework:
The stimulus content dimension
Ages 6-19
Another perspective:

BIS framework:
The stimulus content dimension

Ages 6-19
WJ IV test 2D MDS (Ages 6 to 19; n = 4,082)
BIS content/stimulus characteristic

NmSeries

H (RQ) #

.93 .63 H .80

.63 .73 .64
The BIS intelligence framework (2016)

<table>
<thead>
<tr>
<th>Cognitive Operations</th>
<th>Verbal</th>
<th>Auditory</th>
<th>Figural</th>
<th>Numeric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasoning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creativity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stimulus Content

- **Gf**
  - .78
  - NumSer: .93, .63, H, .80
  - H (RQ): .63, .73, .64

- **ConFrM**
  - .93, .65, M, .66
  - H (I): .44, .47, .35

- **AnlSyn**
  - .92, .62, M, .62
  - H (RG): .25, .43, .34
How to evaluate the unusualness (base rate) of WJ IV cluster or test score differences

Kevin McGrew, PhD.
Educational/School Psychologist
Director
Institute for Applied Psychometrics (IAP)

It’s a pleasure when you use the correct measure

© Institute for Applied Psychometrics; Kevin McGrew 11-23-15
Select WJ IV COG cluster/test score significance values (ages 6-19) *

**Correlation**

- **SD(diff) 1.50** (≈ 13% base rate)
- **SD(diff) 1.65** (≈ 10% base rate)

**GIA (7 tests)**
- SAPT’s (4 tests)
  - .87/≈12/≈13
- Gf-Gc (4 tests)
  - .86/≈12/≈13
- BIA (2 tests)
  - .94/≈8/≈9

**Gc**
- Oral Vocabulary  .71/≈18/≈20
- General Information .97/≈5/≈6

**Gf**
- Number Series  .47/≈24/≈27
- Concept Formation .94/≈8/≈9

**Gwm**
- Verbal Attention .47/≈24/≈27
- Number Reversed .94/≈8/≈9

**Glv**
- Story Recall .34/≈27/≈30

**Gv**
- Visualization .43/≈25/≈28

**Ga**
- Picture Recognition .37/≈27/≈29
- Nonword Repetition .60/≈21/≈24

**Gs**
- Let-Pattern Matching .60/≈21/≈24

* Rounded values calculated in WJ IV norm data (ages 6 to 19)

© Institute for Applied Psychometrics; Kevin McGrew 11-23-15
Intelligence is Important, Intelligence is Complex

Timothy Z. Keith
Alfred University

...and.....”intelligent” intelligent testing is complex....and important
We are the instrument !!!!

In the remainder of this presentation I will model “intelligent” intelligence test interpretation for the WJ IV COG+OL

Will provide you with some aids and templates to organize thinking and test data

Not to be used as cookbooks