Black Box Warning

“cognitive profile analysis in all of its forms has long been implicated as psychometrically flawed (Canivez, 2013; Kranzler et al., 2016; Macmomm & Barnett, 1997; McDermott et al., 1990, 1992; Watkins, 2000) and the ecological/treatment utility of the strengths and weaknesses generated from these analyses has historically been poor (Burns et al., 2016; Elliott & Resing, 2015; Fletcher & Miciak, 2016). As a result, the scores associated with these analyses and the methodologies where employed should be interpreted with caution.”
Overview and Goals

- Discuss historical roots of profile analysis methods.
- Core assumptions of modern profile analytic methods.
- Potential psychometric and conceptual issues.
- Current status of profile analysis in school psychology training and practice.
- Concluding thoughts.
- Q&A.

Some Caveats...

- These are value laden issues.
- Stipulate:
  - Profile analysis has long made intuitive sense.
  - We were all taught to do this stuff too.

Guiding Principle 1: Sources of Evidence

- Evidentiary sources are not equal.
- Sometimes they should not even be regarded as evidence.

- Things to consider:
  - Quality of
  - Empirical
  - Peer review
  - Are claims supported
  - Potential conflicts of interest

  Important to consider quality of sources when evaluating the evidence base for a claim.
Guiding Principle 2: Measurement Matters

- Reliability (Consistency in scores/measurement).
- Validity (Appropriateness of inferences based on test scores).
- Diagnostic Utility (Individual applications: prediction for the individual, diagnostic utility, classification of individuals).
- Treatment Validity (Link between a test score or score comparison that reliably indicates some specific treatment or differential application produces differential improvement.

Procedures used for diagnostic decision-making in school psychology must be supported with appropriate scientific evidence above and beyond a subjective belief that such procedures are valid (Lilienfeld, Ammirati, & David, 2012).

History of Intelligence Test Interpretation (Kamphaus, Winsor, Rowe, & Kim, 2005; Kranzler & Floyd, 2013)

First Wave (1900-1940)
- Clinical influence of the Wechsler Scales.
- Early speculation about pathognomonic meaning of subtest profiles.
- Abnormal test scatter (Harris & Shakow, 1938).
- Diagnostic psychological testing (Rappaport, Gil, & Shaffer, 1946).
- First formal system that could be used across tests.
- Systematic step-by-step level of analysis.
- Visual inspection of scores.
- Ipsative assessment (Davis, 1959)
  - Deviation of score from reference anchor
  - plus = strength, minus = weakness

Second Wave (1945-1960s)
- First formal system that could be used across tests.
- Systematic step-by-step level of analysis.
- Visual inspection of scores.
- Ipsative assessment (Davis, 1959)
  - Deviation of score from reference anchor
  - plus = strength, minus = weakness

Third Wave (1960-1970's)
- First formal system that could be used across tests.
- Systematic step-by-step level of analysis.
- Visual inspection of scores.
- Ipsative assessment (Davis, 1959)
  - Deviation of score from reference anchor
  - plus = strength, minus = weakness

Fourth Wave (1980 to Present)
- First formal system that could be used across tests.
- Systematic step-by-step level of analysis.
- Visual inspection of scores.
- Ipsative assessment (Davis, 1959)
  - Deviation of score from reference anchor
  - plus = strength, minus = weakness

Profile Analysis 1.0 (1930-1997)

- Clinical influence of the Wechsler Scales.
- Early speculation about pathognomonic meaning of subtest profiles.
- Abnormal test scatter (Harris & Shakow, 1938).
- Diagnostic psychological testing (Rappaport, Gil, & Shaffer, 1946).
- First formal system that could be used across tests.
- Systematic step-by-step level of analysis.
- Visual inspection of scores.
- Ipsative assessment (Davis, 1959)
  - Deviation of score from reference anchor
  - plus = strength, minus = weakness
Intelligent Testing (IT; Kaufman, 1979; 1993)

- Step 1: Interpret the global composite.
- Step 2: Interpret the index scores.
- Step 3: Interpret subtest scores.
- Step 4: Evaluate index-level strengths and weaknesses.
- Step 5: Evaluate subtest-level strengths and weaknesses.
- Step 6: Evaluate test-session behaviors.

Key Elements of 1.0 Approaches

- Encourage subtest-level analysis.
- Largely test specific (i.e., Wechsler Scales).
- De-emphasize interpretation of FSIQ.
- Scatter analyses.
- Heavy emphasis on ipsative analyses.
- Many of these approaches emerged during a time in which the tools for advanced psychometric analyses were largely underdeveloped.
  - Advanced and maintained largely through clinical lore.

Seminal 1.0 Critiques

- "Just say no" (McDermott et al., 1990)
- Illusions of meaning (McDermott et al., 1992).
- Myth of the master detective (Macmann & Barnett, 1997).
- Generalizability of test session BX (Glutting et al., 1996).
- Why does Wechsler subtest profiles analysis persist (Bray, Kehle, & Hintze, 1998)?
- Shared professional myth (Watkins, 2000).

"scientific psychological practice cannot be sustained by clinical conjectures and personal anecdotes that have consistently failed empirical validation. Consequently, psychologists should eschew interpretation of cognitive test profiles and must accept that they are acting in opposition to the scientific evidence if they engage in this practice."

-Watkins (2000, p. 476)
Long-Term Stability of S/Ws (Watkins & Canivez, 2004)

- 0.01-0.20 slight
- 0.21-0.40 fair
- 0.41-0.60 moderate
- 0.61-0.80 substantial
- 0.81-1.00 almost perfect

Profile Analysis 2.0 (1997-Present)

- Modified IT (Flanagan & Alfonso, 2016)
  - De-emphasize subtest and most ipsative comparisons
  - Scatter analysis
- CHC-inspired approaches to test interpretation:
  - Cross-battery assessment.
  - Dual/discrepancy consistency model.
  - C-LIM.
- Other within and between test approaches:
  - SNP
  - Other PSW methods (e.g., Naglieri, 2011; Hale & Fiorello, 2004).

Assumptions of 2.0 Approaches

- Theoretical advances (i.e., CHC) enable us to make more defensible inferences from score profiles.
- Previous limitations are no longer relevant in most circumstances
- Primary interpretation at the broad ability factor/composite-level
  - The scores reflect legitimate psychological dimensions.
  - If located, these constructs are measured well.
  - Unique profiles and signs are diagnostically valid and have implications for treatment.
  - Score profiles are stable over time.
- FSIQ continues to be de-emphasized.
- Clinicians are able to engage in complex configural analyses (McFall, 1989).
Factor Extraction Criteria for Ability Measures

<table>
<thead>
<tr>
<th>Test</th>
<th>Publication Year</th>
<th>Theory</th>
<th>EFA/PCA</th>
<th>CFA Factors</th>
<th>HIP MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>WJ-III</td>
<td>2001</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>7</td>
</tr>
<tr>
<td>SB-N</td>
<td>2003</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>5</td>
</tr>
<tr>
<td>KABC-II</td>
<td>2004</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>5</td>
</tr>
<tr>
<td>DAS-II</td>
<td>2007</td>
<td>X</td>
<td>X</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>WISC-V</td>
<td>2014</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>5</td>
</tr>
</tbody>
</table>

Note: HIP and MAP values calculated from independent EFA studies.

Applied EFA Example Featuring the WJ-IV
(Dombrowski et al., 2018)

Table 1: Summary of WJ-IV Subtest Variance According to Seidman-Leona Orthogonalization of Four Factors (Ages 9–15)

<table>
<thead>
<tr>
<th>Subtest</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDLX</td>
<td>0.2</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.05</td>
<td>-0.1</td>
</tr>
<tr>
<td>LDLX</td>
<td>0.2</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.05</td>
<td>-0.1</td>
</tr>
<tr>
<td>LDLX</td>
<td>0.2</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.05</td>
<td>-0.1</td>
</tr>
<tr>
<td>LDLX</td>
<td>0.2</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.05</td>
<td>-0.1</td>
</tr>
<tr>
<td>LDLX</td>
<td>0.2</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.05</td>
<td>-0.1</td>
</tr>
</tbody>
</table>

Note: LDLX = Letter-Number Span X; AUD = auditory memory, AND = auditory processing, OR = oral spatial spanning, LV = language-verbal; C = comprehensive knowledge; G = general knowledge; PD = processing speed; G = g-true; G = g-false. Loadings ≥ 0.3 are marked. Dombrowski et al. (2018).}

CFA Replication (Dombrowski et al., 2018)

Table 1: CFA Fit Statistics for the Age 9 to 15 Group

<table>
<thead>
<tr>
<th>Model</th>
<th>CMIN/DF</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>GFI</th>
<th>AGFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: HDLX</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.85</td>
<td>0.82</td>
</tr>
<tr>
<td>Model 2: HDLX</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.85</td>
<td>0.82</td>
</tr>
<tr>
<td>Model 3: HDLX</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.85</td>
<td>0.82</td>
</tr>
<tr>
<td>Model 4: HDLX</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.85</td>
<td>0.82</td>
</tr>
<tr>
<td>Model 5: HDLX</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.85</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Another Potential Confound (Carroll, 1993; 1995)

Table 1: Omega Estimates for Contemporary Cognitive Measures Based Upon Suggested Alignment with CHC Theory

<table>
<thead>
<tr>
<th>Test</th>
<th>Source/Method</th>
<th>Subtests</th>
<th>g</th>
<th>Gc</th>
<th>Gf</th>
<th>Ga</th>
<th>Gv</th>
<th>Gsm/wm</th>
<th>Glr</th>
<th>Gs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAS II</td>
<td>Canivez &amp; McGill (2016)</td>
<td>EFA</td>
<td>.819</td>
<td>.266</td>
<td>.066</td>
<td>.140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KABC II</td>
<td>McGill &amp; Dombrowski (2017)</td>
<td>EFA</td>
<td>.800</td>
<td>.197</td>
<td>*</td>
<td>*</td>
<td>.393</td>
<td>.236</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAIS IV</td>
<td>Gignac &amp; Watkins (2013)</td>
<td>CFA</td>
<td>.840</td>
<td>.290</td>
<td>*</td>
<td>*</td>
<td>.280</td>
<td>.390</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WISC IV Spanish</td>
<td>McGill &amp; Canivez (2016)</td>
<td>EFA</td>
<td>.833</td>
<td>.280</td>
<td>*</td>
<td>*</td>
<td>.120</td>
<td>.179</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WISC V</td>
<td>Dombrowski et al. (2015)</td>
<td>EFA</td>
<td>.849</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>.190</td>
<td>.162</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CFA</td>
<td>.829</td>
<td>.373</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stability of Composite Scores

- Long-term stability of scores and unique PSWs (e.g., Watkins & Smith, 2013).
  - ~30-40% of participants' index scores fluctuated by >10 points.
- Examiner bias (McDermott, Watkins, & Rhoad, 2014).
  - All scores suffered from non-trivial assessor bias.
  - Implications for PSW?

Diagnostic Utility of PSW

- These findings have been replicated in every subsequent study examining DW of various PSW methods and permutations (Brandler et al., 2016, 2019; McGill, Conoyer, & Feller, 2018; Miciak et al., 2016; Miciak, Taylor, Denton, & Fletcher, 2015; Taylor, Miciak, Fletcher, & Frances 2017).
  - No comparable study reporting positive findings has been produced.
  - Factors that degrade decision utility:
    - Test selection.
    - More assessment.
      - Dilution effect (Faust, 1989; Nisbett et al., 1981)

Evidence-Based?

“Most defenders of the role of cognitive assessment in the diagnosis and treatment of learning disabilities believe that a comprehensive cognitive assessment is essential and that knowing about a person’s cognitive abilities leads to better decisions about interventions (Hale et al., 2010). After rereading dozens of papers defending such assertions, including our own, we can say that this position is mostly backed by rhetoric in which assertions are backed by citations of other scholars making assertions backed by citations of still other scholars making assertions. There is nothing wrong with making and citing assertions, but to confuse such assertions with epistemological bedrock is to be epistemologically confused and groundless.”

Evidence-Based?

“Most defenders of the role of cognitive assessment in the diagnosis and treatment of learning disabilities believe that a comprehensive cognitive assessment is essential and that knowing about a person’s cognitive abilities leads to better decisions about interventions (Hale et al., 2010). After rereading dozens of papers defending such assertions, including our own, we can say that this position is mostly backed by rhetoric in which assertions are backed by citations of other scholars making assertions backed by citations of still other scholars making assertions. There is nothing wrong with making and citing assertions, but to confuse such assertions with epistemological bedrock is to be epistemologically confused and groundless.


Treatment Utility: Can 200 Studies Be Wrong?

How Popular is Profile Analysis?

• Contemporary surveys continue to indicate that it remains extremely popular among practitioners (Benson et al., 2018; Sotelo & Dynega Dixon, 2014).
  • ~50-70% of respondents report using index-level profile analysis on a routine basis.
  • Over 23% indicate that they continue to interpret subtest-level profiles.
  • Over half indicate that they rely exclusively on the levels-of-analysis approach championed by Kaufman and Sattler.
  • 44% indicate that they never interpret the FSIQ when there is significant scatter.
  • Majority of training programs emphasize intelligent testing and/or XBA (Lockwood & Farmer, 2019).
  • Rise of PSW.
Concluding Thoughts

• Despite being characterized as new and modern, 2.0 profile analysis methods are in many ways simply re-parameterizations of previous practices.
  • Same and in some cases new psychometric issues apply.
  • The Voorhees effect.
  • What can you say versus what can you say with confidence?
• Costs associated with these methods (Williams & Miciak, 2018).
• Implications for informed consent (Posada, 2004).
  "He that forsakes measure, measure forsakes him"
  -Scottish Proverb

A Thought Experiment…

"I ask myself with humility, if my daughter begins struggling with reading, mathematics, or writing...what questions would I want a school psychologist to ask about her?...What does her profile of CHC broad cognitive abilities look like?...Or would I prefer that the school psychologist determine the existence of a perceived problem, complete an ecologically minded assessment and develop low-inference hypotheses to explain the reason for the problem, draw on empirically based interventions to remedy the problem, and collect data to determine whether the interventions led to reductions in the problem? There is no doubt that I would prefer the second option."

- Floyd (2010, pp. 62-63)

Where Do You Go From Here?
Contact Information:
Ryan J. McGill, Ph.D., BCBA-D, NCSP
Assistant Professor of School Psychology
William & Mary School of Education
Professional Website: www.rjmcgill.com
E-Mail: rmcgill@wm.edu

Ryan L. Farmer, Ph.D., BCBA
Assistant Professor of School Psychology
Oklahoma State University
E-Mail: r.farmer@okstate.edu