Advances in Pediatric Neuropsychology Test Interpretation: Importance of Considering Normal Variability and Performance Validity

Brian Brooks, PhD & Michael Kirkwood, PhD, ABPP/CN

Children's Hospital Colorado

Team Collaborators
• David Baker, PsyD
• Amy Connery, PsyD, ABPP
• Robin Peterson, PhD, ABPP
• Tass Simpson, PsyD
• Mike Oshami, MD
• Amy Kanaha, MD
• Jackie Murray, RN
• Wendy Pierde, MD
• Aaron Provance, MD
• Anne Stratton, MD
• Pam Wilson, MD

Student Collaborators
• Robert Blaha, MA
• Cassie Green, MA
• David Hargrave, PsyD
• Christa Hutaff-Lee, PhD
• Jennifer Larson, PhD

External Collaborators/Influences
• Brian Brooks, PhD
• Mike Chafetz, PhD, ABPP
• Jacobus Donders, PhD, ABPP
• John Kirk, PsyD, ABPP
• Bill MacAlister, PhD, ABPP/CN
• Chris Randolph, PhD, ABPP/CN
• Elisabeth Sherman, PhD
• Keith Yeake, PhD, ABPP/CN

Acknowledgements

Financial Disclosure
• No relevant conflicts of interest
  • Employment: Children's Hospital Colorado
  • Consulting: No financial payment for TBI roles. Have served on CDC Pediatric Mild TBI Expert Panel and the Mild TBI/Concussion Work Group, NINDS Common Data Elements Project. Have received payment for consulting work on several non-TBI related projects.
  • Consulting: Financial payment from PAR as expert reviewer during development of the Memory Validity Profile (MVP; Sherman & Brooks, 2016) and Children and Adolescent Memory Profile (CAMP; Sherman & Brooks, 2015). No financial compensation from the sale of the tests.
  • Stock ownership: No relevant investments.
  • Research support: No relevant support.
  • Royalties: No relevant royalties on the sale of these books:

• Financial disclosure
  • Royalties: I receive royalties on a book that I edited on pediatric validity testing.

I. Introduction
1. A Route for Performance Validity Testing in Child and Adolescent Assessment, Michael W. Kirkwood
2. Terminology and Diagnostic Concepts, Elizabeth M. S. Sherman
3. Understanding Deviation from a Developmental Perspective, Eric Potts
4. Performance and Symptom Validity: A Perspective from the Adult Literature
II. Detection Methods and Other Validity Test Usage Matters
5. Review of Pediatric Performance and Symptom Validity Tests, Michael W. Kirkwood
6. Clinical Strategies to Assess the Credibility of Presentations in Children, David A. Baker & Michael W. Kirkwood
7. Motivations Behind Neuropsychological Presentations: Why Children Fudge and How
8. Managing Noncredible Performance in Pediatric Clinical Assessment, Amy K. Connery & Etana Erzy
9. Ethical Considerations in Pediatric Validity Testing, William S. Mandell & Michael W. Kirkwood
III. Validity Testing across Evaluative Settings
10. Child and Adolescent Psychosocial Evaluations, Alison G. Harrison
11. Pediatric Clinical Neuropsychological Evaluations with Medical Populations, Brian L. Biddle
13. Pediatric Forensic Neuropsychological Evaluations, Jacobus Donders

In my very biased opinion, excellent group of authors and recent and relevant state of the science information so...
Historically, reliance on subjective judgment to determine validity in pediatric evaluations

- "Mary appeared to put forth her best effort on all tasks. The results are therefore considered a reliable and valid representation of her cognitive functioning."

Objective instrumentation has allowed us to move away from subjective judgments in vast majority of other domains (e.g., attention, language, memory, mood). Why should test effort be different?

Imagine with intelligence….

- "Mary appeared to have below average intelligence. The results therefore indicate that she has an intellectual disability (aka, mental retardation)."

Problems with relying only on subjective judgment to identify noncredible data

- General literature suggests flaws in clinical judgment and decision-making
- Two neuropsychologically-focused studies by Faust in 1988 (children and adolescents)
  - Youth (9-12, 15-17) told to perform less well than usual but not so obvious that the person testing them would know they were faking
  - No instruction in how to fake
  - Clinicians sent vignette that youth in MVC with LOC, unremarkable CT, and memory complaints some months later; clinicians asked to judge whether data abnormal and then speak to etiology
  - Majority of clinicians thought the profile reflected abnormality
  - Detection rate for malingering 0%
  - Majority of clinicians confident in their judgments
- Faust studies criticized (eg, clinicians have access to more than simply test scores)
  - Bigler (1990); McCaffrey & Lynch (1992)
  - Yel, collectively, raise a number of questions
- Objective methodology has clear potential of reducing classification errors
  - In our experienced group in Denver, many cases would not be identified without PVTs

Consensus Need for Objective Methodology

Independent Evaluations

- NAN (2005)
  - "Symptom exaggeration or fabrication occurs in a sizable minority of neuropsychological examinees, with greater prevalence in forensic contexts. Adequate assessment of response validity is essential in order to maximize confidence both in the results of ability measures and in the diagnoses and recommendations that are based on the results."
- AACN (2009)
  - "Especially because research has shown repeatedly that experienced experts are inaccurate in identifying valid versus invalid ability performances from mere observation of behavior or test scores, for a clinician to choose not to use effort tests and embedded validity indicators requires a solid justification, especially within a forensic context."
- Sweet (2009)
  - "In fact, failure to proactively assess for possible malingering in a forensic case is now considered below the standard of acceptable practice…"

Clinical Evaluations

- NAN (2005)
  - "Although the use of SVTs in clinical contexts may not always be indicated…determinations regarding the validity of patient performance are generally aided by the inclusion of SVTs in neuropsychological evaluations."
- AACN (2009)
  - "Even in a routine clinical context, the presence of problematic effort and response bias can potentially invalidate results. The assessment of effort and genuine reporting of symptoms is important in all evaluations."

Table 2 - Practice Characteristics of Respondents

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>% of total respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Clinical Assessments</td>
<td>261</td>
<td></td>
</tr>
<tr>
<td>Age of Patients Seen for Assessments</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Test batteries used to assess validity</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Language for Assessments</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Clinical interview vs. test scores</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Verbal vs. nonverbal test scores</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Administration settings</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Professional Settings</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Outcome of assessment</td>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 - Methods Used to Detect Invalid Data in Those Under 18 Years of Age

<table>
<thead>
<tr>
<th>Method (Direct Observations of Popularity)</th>
<th>Yes, I use this method (%)</th>
<th>No, I do not use this method (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral observations indicative of poor compliance</td>
<td>92.0</td>
<td>7.1</td>
</tr>
<tr>
<td>Discouragement among rewards</td>
<td>90.8</td>
<td>9.2</td>
</tr>
<tr>
<td>Severity of cognitive impairment with the condition</td>
<td>83.0</td>
<td>17.0</td>
</tr>
<tr>
<td>Patterns of cognitive impairment inconsistent with condition</td>
<td>81.9</td>
<td>18.1</td>
</tr>
<tr>
<td>Impulsive self-reported symptoms to interfere</td>
<td>79.4</td>
<td>20.6</td>
</tr>
<tr>
<td>Hugged validity scale in objective personality or behavioral measures</td>
<td>73.7</td>
<td>26.3</td>
</tr>
<tr>
<td>Scores below expected cutoffs on standardized measures of validity</td>
<td>71.4</td>
<td>28.6</td>
</tr>
<tr>
<td>Scores below cutoff on forced choice test</td>
<td>31.9</td>
<td>28.3</td>
</tr>
<tr>
<td>Implausible changes in test scores</td>
<td>38.3</td>
<td>34.2</td>
</tr>
<tr>
<td>Survey below expected cutoffs on verbalized measures</td>
<td>66.9</td>
<td>30.7</td>
</tr>
<tr>
<td>None</td>
<td>9.7</td>
<td>90.3</td>
</tr>
</tbody>
</table>
The use of formal validity testing as part of the routine assessment of children and adolescents should no longer be considered optional, as it is in alignment with the professional guidelines of the field (e.g., NAN, AACN) and consistent with the ethical guidelines for psychologists (APA, 2002). Integration of performance validity data into neuropsychological practice reflects the current state of the field.


Objective Methods to Evaluate Validity

**PVTs**
- Stand-alone performance-based validity tests
  - Both forced and non-forced choice tests
  - Pros: designed specifically to maximize discriminability between groups so should have better classification statistics
  - Cons: battery time and money
- Indices from conventional tests (“embedded” indicators)
  - Simple cut-offs and atypical performance patterns
  - Pros: time and effort efficient, resistant to coaching, allow for more continuous monitoring of effort
  - Cons: classification statistics generally not as good as stand-alone tests

**SVTs**
- General behavioral/personality inventories
- Disorder-specific inventories

<table>
<thead>
<tr>
<th>PVT</th>
<th>Normal (%)</th>
<th>Random (%)</th>
<th>Sometimes (%)</th>
<th>Often (%)</th>
<th>Almost Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-Item Test</td>
<td>93</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Amsterdam Serial Memory Test</td>
<td>93.5</td>
<td>8.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Annotated Sequencing Test</td>
<td>96.2</td>
<td>7.8</td>
<td>2.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>The B Test</td>
<td>92.0</td>
<td>4.0</td>
<td>3.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CARS</td>
<td>95.3</td>
<td>2.2</td>
<td>1.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CVLT Delayed Recall</td>
<td>77.0</td>
<td>38</td>
<td>20.7</td>
<td>20.7</td>
<td>15.9</td>
</tr>
<tr>
<td>CVLT III Profile</td>
<td>86.6</td>
<td>4.0</td>
<td>3.8</td>
<td>1.0</td>
<td>0</td>
</tr>
<tr>
<td>CVLT III Forced Choice</td>
<td>27.2</td>
<td>7.2</td>
<td>29.9</td>
<td>21.4</td>
<td>15.2</td>
</tr>
<tr>
<td>Dot Counting Test</td>
<td>87.0</td>
<td>3.0</td>
<td>0.5</td>
<td>7.0</td>
<td>0</td>
</tr>
<tr>
<td>MVT</td>
<td>62.0</td>
<td>6.9</td>
<td>14.1</td>
<td>10.9</td>
<td>6.2</td>
</tr>
<tr>
<td>NV-MBT</td>
<td>58.1</td>
<td>5.1</td>
<td>3.3</td>
<td>5.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Relatable Digit Span</td>
<td>34.8</td>
<td>8.0</td>
<td>13.8</td>
<td>27.1</td>
<td>21.0</td>
</tr>
<tr>
<td>ROCF Color Trails</td>
<td>23.1</td>
<td>12.9</td>
<td>31.2</td>
<td>20.7</td>
<td>14.8</td>
</tr>
<tr>
<td>WMTM</td>
<td>50.7</td>
<td>7.2</td>
<td>0.7</td>
<td>1.4</td>
<td>0.0</td>
</tr>
<tr>
<td>Word Completion Memory Test</td>
<td>49.6</td>
<td>7.0</td>
<td>0.0</td>
<td>8.7</td>
<td>5.8</td>
</tr>
<tr>
<td>VNT</td>
<td>85.5</td>
<td>6.2</td>
<td>4.9</td>
<td>2.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Test of Memory Malingering (TOMM)

**What is it?**
- Developed by Tombaugh (1996)
- Examinee presented 50 line drawings twice; forced choice response during IR and DR, with optional retention trial

**Additional tests with potential utility but need more study**
- Several PVTs have been investigated in only one identified pediatric study or by one group
  - Of these:
    - **Most Promising**
      - Nonverbal Medical Symptom Validity Test (Green, 2008)
      - Green, Flaro, Brockhaus, & Montijo (2012); Harrison et al. (2014)
      - Victoria Symptom Validity Test (Slick, Hopp, Strauss, & Thompson, 1997)
      - Brooks (2012)
    - **Mixed Results or Very Little Work**
      - Dot Counting Test (Locke, 1993; Rey, 1964)
      - 21-Item Test (Iverson, 1998)
      - Computerized Assessment of Response Bias (Allen, Conder, Green & Cox, 1997)
      - Courtney, DeRienzo, Allen, & Kuroki (2003); Harrison et al. (2014)
      - Amsterdam Short-Term Memory Test (Schmand & Lindelobus, 2004)
      - Warriner, Swan, & Schneidt (2010)
    - Word Completion Memory Test (WCMT; Hilsabeck & LeCompte, 1997)
      - Warriner, Swan, & Schneidt (2010)

**Stand-Alone PVTs Investigated in Pediatric Populations**
TOMM bottom line
- Most empirical work
- Likely appropriate with children 5+ years
- Appears specific in all but the most impaired children
- Relatively low cost
- Unlikely to be as sensitive as some other measures
(Blaskewitz et al.; Rambo et al., missed 1/3 simulators)
- More time consuming than some other PVTs


Green’s Word Memory Test (WMT) & Medical Symptom Validity Test (MSVT)

- What are they?
  - WMT: Patient presented twice with 20 semantically linked words on computer
  - MSVT: Patient presented twice with 10 semantically linked words
- Followed by a number of trials
- Primary effort measures: IR, DR, and Consistency between two trials
- Originally normed for adults but Flaro provided data from children with variety of clinical disorders
- Profile analysis allows for examination of whether a fail is a “true impairment profile”

Boat
Water
Dog
Cat

Green’s Word Memory Test (WMT)

WMT

<table>
<thead>
<tr>
<th>Source</th>
<th>Population</th>
<th>N</th>
<th>Age Range</th>
<th>Mean (SD)</th>
<th>IR % (SD)</th>
<th>DR % (SD)</th>
<th>Consistency % (SD)</th>
<th>% Rejection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blaskewitz et al. (2015)</td>
<td>Moderate to Severe Emotional Disturbance</td>
<td>18</td>
<td>9 - 12</td>
<td>8.1 (2.3)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>Blaskewitz et al. (2015)</td>
<td>Pediatric Clinical sample</td>
<td>100</td>
<td>-</td>
<td>7.1 (2.3)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>Canavan et al. (2014)</td>
<td>U.S. Clinical sample - younger group</td>
<td>56</td>
<td>10 - 17</td>
<td>11.4 (2.5)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99%</td>
</tr>
<tr>
<td>Canavan et al. (2014)</td>
<td>U.S. Clinical sample - older group</td>
<td>55</td>
<td>6 - 9</td>
<td>8.1 (2.2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99%</td>
</tr>
<tr>
<td>Gooden et al. (2015)</td>
<td>Available (Revised - 2015)</td>
<td>68</td>
<td>11 - 14</td>
<td>11.2 (2.0)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99%</td>
</tr>
<tr>
<td>Glass et al. (2015)</td>
<td>Available (Revised - 2015)</td>
<td>55</td>
<td>6 - 11</td>
<td>8.4 (1.8)</td>
<td>80.6 (7.8)</td>
<td>83.5 (9.3)</td>
<td>-</td>
<td>99%</td>
</tr>
</tbody>
</table>

WMT & MSVT Bottom Line
- Good evidence appropriate with children with 3rd grade reading level or better
  - Solid specificity above this threshold
- Evidence to indicate more sensitive than TOMM (Blaskewitz et al.; Rambo et al., )
  - Consistent with our experience in Denver
- Available multiple languages
- Potential added benefit of “profile analysis” to detect true impairment vs. noncredible effort
- MSVT can be administered quickly so good as screening measure
- MSVT cost per use; WMT annual fee (historically)

MSVT

<table>
<thead>
<tr>
<th>Source</th>
<th>Population</th>
<th>N</th>
<th>Age Range</th>
<th>Mean (SD)</th>
<th>IR % (SD)</th>
<th>DR % (SD)</th>
<th>Consistency % (SD)</th>
<th>% Rejection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blaskewitz et al. (2015)</td>
<td>Wired Emotionally Disturbed</td>
<td>34</td>
<td>11 - 13</td>
<td>13.0 (3.1)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99%</td>
</tr>
<tr>
<td>Blaskewitz et al. (2015)</td>
<td>U.S. Clinical sample - younger group</td>
<td>56</td>
<td>10 - 17</td>
<td>11.4 (2.5)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99%</td>
</tr>
<tr>
<td>Canavan et al. (2014)</td>
<td>U.S. Clinical sample - older group</td>
<td>55</td>
<td>6 - 9</td>
<td>8.1 (2.2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99%</td>
</tr>
<tr>
<td>Canavan et al. (2014)</td>
<td>U.S. Clinical sample - younger group</td>
<td>56</td>
<td>10 - 17</td>
<td>11.4 (2.5)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99%</td>
</tr>
<tr>
<td>O’Connor et al. (2015)</td>
<td>Wired Emotionally Disturbed</td>
<td>38</td>
<td>-</td>
<td>11.4 (2.9)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99%</td>
</tr>
<tr>
<td>Kirkwood &amp; Alt (2015)</td>
<td>Wired Emotionally Disturbed</td>
<td>193</td>
<td>8 - 11</td>
<td>9.2 (2.5)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99%</td>
</tr>
<tr>
<td>Sh功力 et al. (2015)</td>
<td>U.S. Special Education</td>
<td>25</td>
<td>6 - 10</td>
<td>9.1 (2.4)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99%</td>
</tr>
<tr>
<td>Blaskewitz et al. (2015)</td>
<td>Wired Emotionally Disturbed</td>
<td>55</td>
<td>6 - 11</td>
<td>8.4 (1.8)</td>
<td>80.6 (7.8)</td>
<td>83.5 (9.3)</td>
<td>-</td>
<td>99%</td>
</tr>
</tbody>
</table>

Rey Fifteen-Item Test (FIT)

- What is it?
  - Best known of Rey’s validity procedures
  - Patient shown 15 items and then asked to draw as many as can
  - Adapted by others – eg, Boone et al. (2002) developed a recognition format

A B C
1 2 3
a b c
i j k

Embedded Indicators

- Extensive literature in adult populations (Boone, 2007; Larrabee, 2007)
- Much less study in children

Digit Span as an Embedded Indicator

- Adult studies
  - Dozens of studies across a variety of samples (Reviews: Suhr & Barrash, 2007; Babikian & Boone, 2007)
  - Age-corrected scaled scores
  - Reliable Digit Span (Cronbach, Baker, & Gola, 1994)
    - Calculated by summing the longest string of digits repeated without error over two trials under both forward and backward conditions
    - Eg, pass both trials ≤ 3 digits forward, pass both trials of two digits back ≤ 5
    - Cutoff of 8 to 10: these produce sensitivity values above 95% in nearly all adult studies (specificity less ideal in more severely affected populations at this level)
    - Cutoff ≤ 6: sensitivity is lowered to around 40-60% but specificity improves more consistently to at least 90%

- First child study: Blaskewitz, Merten, & Kathmann (2008)
  - German simulation design with 70 children (6 – 11 year olds)
  - WISC-III Digit Span subtest administered
  - Using adult cutoff for RDS, majority of matched controls (59%) failed
  - Classification statistics for lower RDS cutoff scores and other Digit Span scores not published

Memory Validity Profile (Sherman & Brooks – 2016)

- First commercially available stand-alone PVT designed specifically for children/teens
- Exciting development for all of us pediatric neuropsychologists
- Consists of verbal and visual paradigms
- Underwent test development like commercially produced cognitive tests including pilot testing, expert panel review, bias review, and refinement testing
- Normalized on 1,200 US youth aged 5-21 years, 200 youth with clinical diagnoses, and 45 children in a simulation design study
- First validity test with age-adjusted cut scores to minimize false positives in young children
- Not yet available for independent review but certainly promising

Embedded PVTs Investigated in Pediatric Populations

- Review

- Digit Span
  - Blaskewitz et al. (2008)
  - Kirkwood et al. (2011)
  - Weishon et al. (2012)
  - Loughan et al. (2012)
  - Perna et al. (2014)
  - Harrison & Armstrong (2014)

- CVLT-C
  - Baker et al. (2014)
  - Brooks et al. (2015)

- Automated Sequences Task
  - Kirkwood et al. (2014)

- Matrix Reasoning
  - Kirkwood et al. (2012)
  - Rambo et al. (2013)

- Symptom Validity Scale
  - Chaheifetz et al. (2007; 2008)
  - CNS Vital Signs
    - Brooks et al. (2014)

- ChAMP
  - Sherman & Brooks (2015)
Denver results compare favorably to many real-world adult populations
- Moderate sensitivity (~50%) when specificity > 90%
- Loughan et al (2012) found similar classification statistics using a cut-score of ss ≤ 4 (Sens = 43%, Spec 91%)
- Only 7 noncredible cases total though; 67 mild TBI cases
- Keep in mind nature of sample
  - Higher functioning older kids/teens with mild neurological injury
- Different results will almost certainly be obtained in lower functioning populations (e.g., those with neurologically or developmentally-based problems)
  - Indeed…

Digit Span as Embedded Indicator

RDS scores showed strong correlations with clinical and cognitive variables, including age of participant and intellectual functioning.
- Overall pass rate of RDS scores at ≤ 6 was low (65%)

Embedded Indicators from CVLT

- Adult studies
  - Dozens of studies across a variety of samples
  - Recognition scores generally most sensitive
- Much less attention in children
  - We’ve looked at in our mild TBI sample
  - Most recently, N = 411 (aged 8-16 yo)

Again…

- Different results will almost certainly be obtained in lower functioning populations (e.g., those with neurologically or developmentally-based problems)
Child & Adolescent Memory Profile
(Sherman & Brooks, 2015)

- First commercially available pediatric test to include embedded indicators
- Brief memory battery with two verbal and two visual subtests
- Subtests contain embedded indicators using three-item forced-choice responding
- Cutoffs based on below chance responding
- 1,200 youth aged 5-21 years, 200 youth with clinical diagnoses, and 45 children in a simulation design study

Table 7 Frequency Use of SVTs with Children and Adolescents.

<table>
<thead>
<tr>
<th>SVT</th>
<th>Never (%)</th>
<th>Rare (%)</th>
<th>Sometimes (%)</th>
<th>Often (%)</th>
<th>Always or Always (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASC-2 Validity Indicator</td>
<td>32.2</td>
<td>6.9</td>
<td>12.7</td>
<td>19.2</td>
<td>10.2</td>
</tr>
<tr>
<td>BRIEF Validity Indicator</td>
<td>27.2</td>
<td>7.2</td>
<td>12.6</td>
<td>21.0</td>
<td>15.8</td>
</tr>
<tr>
<td>MWPS-4A Indicators</td>
<td>47.1</td>
<td>15.6</td>
<td>13.4</td>
<td>12.4</td>
<td>10.6</td>
</tr>
<tr>
<td>Personality Inventory for Youth Validity Indicators</td>
<td>86.8</td>
<td>4.3</td>
<td>3.9</td>
<td>3.2</td>
<td>2.5</td>
</tr>
<tr>
<td>Trauma Symptom Checklist for Children Validity Indicators</td>
<td>87.3</td>
<td>5.8</td>
<td>4.7</td>
<td>6.7</td>
<td>1.4</td>
</tr>
</tbody>
</table>

BRIEF (and other domain-specific scales)

- No identified independent studies examining faking bad or negativity scales

<table>
<thead>
<tr>
<th>BASC-2 SRP Validity Scale</th>
<th>Rare/Unlikely</th>
<th>Within 'Caution' or 'Extreme Caution'</th>
<th>Likely/Probable</th>
</tr>
</thead>
<tbody>
<tr>
<td>F Scale</td>
<td>8</td>
<td>20</td>
<td>95</td>
</tr>
<tr>
<td>L Scale</td>
<td>283</td>
<td>223</td>
<td>40</td>
</tr>
<tr>
<td>V Scale</td>
<td>222</td>
<td>222</td>
<td>40</td>
</tr>
<tr>
<td>Response Pattern</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Consistency Scale</td>
<td>213</td>
<td>213</td>
<td>40</td>
</tr>
<tr>
<td>Any Validity Scale</td>
<td>195</td>
<td>195</td>
<td>40</td>
</tr>
</tbody>
</table>
Denver Mild TBI Case Series & BASC2 Conclusions

- First identified study to examine a self-report validity scale in a real-world pediatric sample of noncredible responders
- Vast majority of patients who failed the MSVT provided valid self-report BASC-2 profiles
- Data contrasts with many adult studies demonstrating self-report validity scales strongly associated with PVT performance
- Sole reliance on validity indicators from the BASC-2 (and other child self-report scales?) likely to substantially underestimate the number of patients providing invalid data during neuropsychological evaluation

Initial Data Regarding Effectiveness of Colorado Feedback Model

- In general, very high rate of satisfaction with neuropsych service in mild TBI clinic (~95%)
- Kirkwood, Peterson, Connelly, & Baker, in submission
- Examination of service in credible vs. noncredible responders (Connery, Peterson, Baker, & Kirkwood, in submission)
- No difference in caregiver satisfaction rates
- Actually see greater symptom reduction in noncredible responders

Table II: Frequency of Statements to Communicate (Verbally or in Report) Noncredible/Invalid Data.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Never (%)</th>
<th>Rarely (%)</th>
<th>Sometimes (%)</th>
<th>Often (%)</th>
<th>Almost Always (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test results are invalid</td>
<td>16.6</td>
<td>23.0</td>
<td>40.4</td>
<td>18.1</td>
<td>7.4</td>
</tr>
<tr>
<td>Test results indicate inadequate effort to perform well</td>
<td>12.1</td>
<td>12.5</td>
<td>35.9</td>
<td>33.6</td>
<td>6.0</td>
</tr>
<tr>
<td>No hit results can be drawn</td>
<td>9.3</td>
<td>14.2</td>
<td>37.4</td>
<td>31.2</td>
<td>6.0</td>
</tr>
<tr>
<td>Test results are inconsistent with severity of condition</td>
<td>5.7</td>
<td>12.8</td>
<td>40.6</td>
<td>25.3</td>
<td>5.7</td>
</tr>
<tr>
<td>Test results indicate inappropriate engagement</td>
<td>15.9</td>
<td>12.9</td>
<td>40.3</td>
<td>23.1</td>
<td>4.5</td>
</tr>
<tr>
<td>Test results indicate poor cooperation</td>
<td>18.9</td>
<td>26.5</td>
<td>37.9</td>
<td>18.2</td>
<td>4.5</td>
</tr>
<tr>
<td>Test results indicate exaggeration or faking</td>
<td>28.5</td>
<td>21.7</td>
<td>32.8</td>
<td>6.9</td>
<td>2.3</td>
</tr>
<tr>
<td>Test results indicate malingering</td>
<td>44.9</td>
<td>28.7</td>
<td>5.7</td>
<td>0.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Note: Data presented in descending order based on “Almost Always”.

Children’s Hospital Colorado Opening Statement for Parent Feedback in Face of Noncredible Effort (Connery, Baker, Peterson, & Kirkwood)

“Whenever we do these evaluations, we give tests that measure whether children are trying their best to do well in order to make sure the test results are valid. In other words, when a child does not do well on testing, we want to make sure that it is due to an actual weakness rather than to a child not trying his/her best. During today’s evaluation, these tests showed that XXX was not always trying his/her best to do well. What are your thoughts about this? Do you have ideas on why this might have happened?”

Rationale for Using PVTs with School-Aged Children and Adolescents

1) Children are capable of deception
2) Noncredible presentations occur consistently in pediatric cognitive assessments
3) We have empirically-backed objective methods to help detect invalid data – why not use?
4) Failure on PVTs has significant implications
   - Data interpretation
   - Clinical management
   - Systematically

### General Pediatric Clinical Case Series

<table>
<thead>
<tr>
<th>Source</th>
<th>Population</th>
<th>N</th>
<th>Age</th>
<th>PVT</th>
<th>% Noncredible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donders (2005)</td>
<td>Mixed Neuro</td>
<td>190</td>
<td>6 – 16</td>
<td>TOMM</td>
<td>2%</td>
</tr>
<tr>
<td>Carone (2008)</td>
<td>Moderate- Severe Brain Injury</td>
<td>38 (mean: 11.8)</td>
<td>MSVT</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>MacAllister, Nakhtina, Bender, Karantzoulis, &amp; Carlson (2009)</td>
<td>Epilepsy</td>
<td>60</td>
<td>6 – 17</td>
<td>TOMM</td>
<td>3%</td>
</tr>
<tr>
<td>Green et al. (2010)</td>
<td>Mixed Neuro/Dev</td>
<td>380</td>
<td>WMT</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Green et al. (2010)</td>
<td>Mixed Neuro/Dev</td>
<td>265</td>
<td>MSVT</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>Kirk, Harris, HualLea, Koelmay, Osinski, &amp; Kirkwood (2011)</td>
<td>Mixed Neuro/Dev</td>
<td>150</td>
<td>5 – 16</td>
<td>TOMM</td>
<td>4%</td>
</tr>
<tr>
<td>Brooks (2012)</td>
<td>Mixed Neuro</td>
<td>150</td>
<td>6 – 19</td>
<td>VSVT</td>
<td>9%</td>
</tr>
<tr>
<td>Povitz, Moskowitz, Kirkwood, Sherman, &amp; Brooks (2014)</td>
<td>Mixed Neuro</td>
<td>286</td>
<td>5 – 18</td>
<td>TOMM</td>
<td>3%</td>
</tr>
</tbody>
</table>

### Pediatric Case Series: Mild TBI

<table>
<thead>
<tr>
<th>Source</th>
<th>Population</th>
<th>N</th>
<th>Age</th>
<th>PVT</th>
<th>% Noncredible Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children’s Hospital Colorado</td>
<td>Mild TBI (clinical)</td>
<td>1000+ total</td>
<td>8 – 17</td>
<td>MSVT + TOMM Rey FIT Various embedded measures</td>
<td>12 – 19%</td>
</tr>
<tr>
<td>Kirkwood &amp; Kirk (2010); Kirkwood et al. (2011); Kirkwood et al. (2012); Kirkwood et al. (2013); Baker et al. (2013); Green et al. (2014); Kirk et al. (2014); Kirkwood et al. (2014); Lenzen et al. (2015)</td>
<td>Mild TBI (clinical)</td>
<td>382</td>
<td>8 – 16</td>
<td>RDS Digit Span</td>
<td>20%</td>
</tr>
</tbody>
</table>

### Pediatric Case Series: Independent Setting (Social Security Disability)

<table>
<thead>
<tr>
<th>Source</th>
<th>Population</th>
<th>N</th>
<th>Age</th>
<th>PVT</th>
<th>% Noncredible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chafetz et al. (2007); Chafetz (2008)</td>
<td>Social Security Disability Claimants (independent)</td>
<td>123</td>
<td>6 – 16</td>
<td>TOMM MSVT</td>
<td>48-60% (25-50% PVT chance level or below)</td>
</tr>
</tbody>
</table>

### Implications of PVT Failure for Interpreting Other Data During a Cognitive Exam

**SO WHAT?**

- Multiple studies with adults have suggested that PVT performance relates strongly to ability-based tests
  - Green et al., 2001; Constantinou et al., 2005; Green, 2007; Lange et al., 2010; Meyer et al., 2011
  - In these samples (mostly compensation-seeking), ~50% variance in neuropsychological test scores explained by PVT performance (much more variance than explained by brain injury severity, education, age, etc.)
  - Up until few years ago, no identified studies in pediatric populations: similar effects?

If MSVT measures effort, not ability, two expectations:

1. MSVT performance should be unrelated to demographic, developmental, and injury-related factors in sample
2. MSVT performance should relate to a wide range of tests across the battery – not just those tests tapping memory or related skills (e.g., reading) that seem necessary on surface to complete the MSVT
Support for idea PVT measures effort rather than ability
- No background or injury-related variable differentiated those who passed from those who failed
- PVT results explained ~40% of the variance across the test battery

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASI</td>
<td>110.8</td>
<td>25.1</td>
<td>102.5</td>
<td>20.5</td>
<td>-3.5</td>
<td>.001</td>
</tr>
<tr>
<td>Trajectory Test</td>
<td>70.8</td>
<td>20.7</td>
<td>62.9</td>
<td>15.8</td>
<td>-3.5</td>
<td>.01</td>
</tr>
<tr>
<td>Visual Reaction Time Test</td>
<td>60.8</td>
<td>15.8</td>
<td>55.9</td>
<td>12.5</td>
<td>-3</td>
<td>.01</td>
</tr>
<tr>
<td>Reading Test</td>
<td>50.9</td>
<td>20.7</td>
<td>45</td>
<td>15.8</td>
<td>-3</td>
<td>.01</td>
</tr>
<tr>
<td>Testimens Test</td>
<td>20.9</td>
<td>5.5</td>
<td>16.9</td>
<td>4.5</td>
<td>-3</td>
<td>.01</td>
</tr>
<tr>
<td>Long Items Data Reduction Test</td>
<td>10.9</td>
<td>3.5</td>
<td>8.9</td>
<td>2.5</td>
<td>-3</td>
<td>.01</td>
</tr>
<tr>
<td>DART Test</td>
<td>80.9</td>
<td>20.7</td>
<td>72.9</td>
<td>15.8</td>
<td>-3</td>
<td>.01</td>
</tr>
<tr>
<td>PASI Score</td>
<td>40.9</td>
<td>15.8</td>
<td>35.9</td>
<td>12.5</td>
<td>-3</td>
<td>.01</td>
</tr>
<tr>
<td>PASI Total Score</td>
<td>30.9</td>
<td>10.7</td>
<td>25.9</td>
<td>10.7</td>
<td>-3</td>
<td>.01</td>
</tr>
<tr>
<td>PASI Total Score</td>
<td>20.9</td>
<td>5.5</td>
<td>16.9</td>
<td>4.5</td>
<td>-3</td>
<td>.01</td>
</tr>
<tr>
<td>PASI Total Score</td>
<td>10.9</td>
<td>3.5</td>
<td>8.9</td>
<td>2.5</td>
<td>-3</td>
<td>.01</td>
</tr>
<tr>
<td>PASI Total Score</td>
<td>5.9</td>
<td>2.5</td>
<td>4.9</td>
<td>1.5</td>
<td>-3</td>
<td>.01</td>
</tr>
<tr>
<td>PASI Total Score</td>
<td>2.9</td>
<td>1.5</td>
<td>1.9</td>
<td>.5</td>
<td>-3</td>
<td>.01</td>
</tr>
</tbody>
</table>

Several studies with adults have also suggested that performance on PVTs has significant effect on postconcussive symptoms report after mild TBI
- Lange et al. (2010)
- Iverson et al. (2010)
- Tsanidis et al. (2008)

No identified studies in pediatric populations: similar effects?
- In Denver mild TBI series, children failing MSVT reported significantly more "postconcussive symptoms" than those who pass MSVT ($p < .001; d = 1.1$
  - Kirkwood, Peterson, Connelly, Baker, & Grubenhoff (2014)

Implications of PVT Failure for Interpreting Symptom Report
- As one example...
- Social Security Administration
  - In 2011, for malingered mental disorders in adults, estimated cost to SSA was $20.02 billion
  - In 2011, for malingered mental disorders in children, estimated cost to SSA was $2.13 billion
- Given that many pediatric providers do not routinely use PVTs, likely an underestimate when collectively consider governmental, legal, healthcare, and educational costs

Implications of PVT Failure for Broader Systems