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

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

Acknowledgements

Children's Hospital Colorado

Team Collaborators
- David Baker, PsyD
- Amy Connery, PsyD, ABPP
- Robin Peterson, PhD, ABPP
- Tess Simpson, PsyD
- Mike Dicicharo, MD
- Amy Kanallakan, MD
- Jackie Murray, RN
- Wendy Pierce, MD
- Aaron Provance, MD
- Anne Stratton, MD
- Pam Wilson, MD

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

External Collaborators/Influences
- Brian Brooks, PhD
- Mike Chafetz, PhD, ABPP
- Jacobus Donders, PhD, ABPP
- John Kirk, PsyD, ABPP
- Bill MacAllister, PhD, ABPP/CN
- Chris Randolph, PhD, ABPP/CN
- Elisabeth Sherman, PhD
- Keith Yeates, PhD, ABPP/CN
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 (ChAMP; Sherman & Brooks, 2015). No financial compensation from the sale of the tests.
  - Stock ownership: No relevant investments.
  - Research support: Principal investigator, co-investigator, or collaborator on pediatric TBI grants funded by numerous organizations including CDC, NIDRR, NIMH, Colorado TBI Trust Fund, Children's Hospital Colorado.
  - Honoraria: Have received honoraria and expense reimbursement from multiple institutions and professional scientific bodies for presenting pediatric TBI and validity testing work at meetings, conferences, and symposiums.
  - Royalties: I receive royalties on the sale of these books:

Financial Disclosure

• Financial disclosure
  - Royalties: I receive royalties on a book that I edited on pediatric validity testing.
In my very biased opinion, excellent group of authors and recent and relevant state of the science information so….

I. Introduction
1. A Rationale for Performance Validity Testing in Child and Adolescent Assessment, Michael W. Kirkwood
2. Terminology and Diagnostic Concepts, Elisabeth M. S. Sherman
3. Understanding Deception from a Developmental Perspective, Eric Peterson & I
4. Performance and Symptom Validity: A Perspective from the Adult Literature, C

II. Detection Methods and Other Validity Test Usage Matters
5. Review of Pediatric Performance and Symptom Validity Tests, Michael W. Kirk
6. Clinical Strategies to Assess the Credibility of Presentations in Children, Domi
8. Managing Noncredible Performance in Pediatric Clinical Assessment, Amy K. Connery & Yana Suchy
9. Ethical Considerations in Pediatric Validity Testing, William S. MacAllister & Mars

III. Validity Testing across Evaluative Settings
10. Child and Adolescent Psychoeducational Evaluations, Allyson G. Harrison
11. Pediatric Clinical Neuropsychological Evaluations with Medical Populations, Bri
12. Pediatric Sports-Related Concussion Evaluations, Martin L. Rohling, Jennifer Lan, Melissa M. Womble
13. Pediatric Forensic Neuropsychological Evaluations, Jacobus Donders

Objective Methods to Detect Noncredible Data

aka, "effort" or "malingering" tests but movement away from these terms

Performance Validity Tests (PVTs)
Used to detect inadequate effort or noncredible performance during testing

Symptom Validity Tests (SVTs)
Used to detect noncredible responding during self-report measures

A survey of neuropsychologists’ use of validity tests with children and adolescents

Brian L. Brooks1,2,3, Danielle M. Ploetz4, and Michael W. Kirslis3,5

1Neurosciences Program (Brain Injury and Rehabilitation), Alberta Children’s Hospital, Calgary, AB, Canada
2Departments of Pediatrics, Clinical Neurosciences, and Psychology, University of Calgary, Calgary, AB, Canada
3Behaviour and the Developing Brain Program, Alberta Children’s Hospital Research Institute, Calgary, AB, Canada
4Children’s Hospital Colorado, Aurora, CO, USA
5Department of Physical Medicine and Rehabilitation, University of Colorado School of Medicine, Aurora, CO, USA

A web-based survey of validity test use by North American neuropsychologists was conducted, with 282 participants meeting inclusion criteria. Respondents indicated that they use a median of one stand-alone performance validity test (PVT), one embedded PVT, and one symptom validity test (SVT) per pediatric assessment. The vast majority of respondents indicated they gave at least one PVT (92%) and at least one SVT (88%) during each pediatric assessment. A meaningful difference in validity use (i.e., at least a medium effect size) was only found for those who engage in forensic work, with those directors giving more stand-alone PVTs than those who do not conduct forensic work.

The most frequently used validity measures in pediatric assessments are presented, as are reasons participants reported for both using and not using validity tests. Limitations and qualitative comparisons to other surveys on validity test use with adults are discussed.

Table 2: Practice Characteristics of Respondents.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>% of total sample that report any activity in this area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Clinical Assessments Per Month</td>
<td>282</td>
<td>(mean = 8.5, SD = 7.6)</td>
</tr>
<tr>
<td>Ages of Patients Seen for Assessments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-3 years</td>
<td>185</td>
<td>65.6</td>
</tr>
<tr>
<td>6-12 years</td>
<td>267</td>
<td>94.7</td>
</tr>
<tr>
<td>13-17 years</td>
<td>277</td>
<td>98.2</td>
</tr>
<tr>
<td>18+ years</td>
<td>233</td>
<td>82.6</td>
</tr>
<tr>
<td>Language for Assessments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English (100% of the time)</td>
<td>238</td>
<td>84.4</td>
</tr>
<tr>
<td>Spanish (At least some of the time)</td>
<td>35</td>
<td>13.3</td>
</tr>
<tr>
<td>French (At least some of the time)</td>
<td>4</td>
<td>1.6</td>
</tr>
<tr>
<td>Other Languages (At least some of the time)</td>
<td>16</td>
<td>5.8</td>
</tr>
<tr>
<td>Professional Settings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private Practice</td>
<td>146</td>
<td>52.1</td>
</tr>
<tr>
<td>Hospital</td>
<td>177</td>
<td>62.0</td>
</tr>
<tr>
<td>Schools</td>
<td>10</td>
<td>3.6</td>
</tr>
<tr>
<td>Prison/Department Centre</td>
<td>2</td>
<td>0.8</td>
</tr>
<tr>
<td>Psychiatric Facility</td>
<td>7</td>
<td>2.8</td>
</tr>
<tr>
<td>Academics</td>
<td>28</td>
<td>9.3</td>
</tr>
<tr>
<td>Other</td>
<td>12</td>
<td>4.3</td>
</tr>
<tr>
<td>Professional Activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Assessment</td>
<td>275</td>
<td>97.9</td>
</tr>
<tr>
<td>Forensic Medico-Legal</td>
<td>89</td>
<td>31.7</td>
</tr>
<tr>
<td>Therapy</td>
<td>72</td>
<td>25.6</td>
</tr>
<tr>
<td>Trauma Supervision</td>
<td>163</td>
<td>58.0</td>
</tr>
<tr>
<td>Research</td>
<td>127</td>
<td>45.2</td>
</tr>
<tr>
<td>Classroom Teaching</td>
<td>39</td>
<td>13.9</td>
</tr>
<tr>
<td>Administration</td>
<td>136</td>
<td>49.3</td>
</tr>
<tr>
<td>Other activity (e.g., didactics)</td>
<td>13</td>
<td>4.6</td>
</tr>
</tbody>
</table>
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)."

### Table 3 Methods Used to Detect Invalid Data in those Under 18 years of Age.

<table>
<thead>
<tr>
<th>Method (in Descending Order 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>92.9</td>
<td>7.1</td>
</tr>
<tr>
<td>Discrepancies among records</td>
<td>90.8</td>
<td>9.2</td>
</tr>
<tr>
<td>Severity of cognitive impairment inconsistent with the condition</td>
<td>83.0</td>
<td>17.0</td>
</tr>
<tr>
<td>Pattern of cognitive impairment inconsistent with condition</td>
<td>81.9</td>
<td>18.1</td>
</tr>
<tr>
<td>Implausible self-reported symptoms in interview</td>
<td>79.4</td>
<td>20.6</td>
</tr>
<tr>
<td>Flagged validity scales in objective personality or behavioral measures</td>
<td>73.7</td>
<td>26.3</td>
</tr>
<tr>
<td>Score below empirical cutoffs on stand-alone measures of validity</td>
<td>73.4</td>
<td>26.6</td>
</tr>
<tr>
<td>Scores below chance on forced choice test</td>
<td>71.9</td>
<td>28.1</td>
</tr>
<tr>
<td>Implausible changes in test scores</td>
<td>65.8</td>
<td>34.2</td>
</tr>
<tr>
<td>Scores below empirical cutoffs on embedded measures</td>
<td>60.3</td>
<td>39.7</td>
</tr>
<tr>
<td>None</td>
<td>0.7</td>
<td>99.3</td>
</tr>
</tbody>
</table>
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 results)
  - Bigler (1990); McCaffrey & Lynch (1992)
  - Yet, 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.”
Bill MacAllister and Marsha Vasserman (2015)

- 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 5: Frequency Use of PVTs in Children and Adolescents

<table>
<thead>
<tr>
<th>PVT</th>
<th>Never (%)</th>
<th>Rarely (%)</th>
<th>Sometimes (%)</th>
<th>Often (%)</th>
<th>Almost Always (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-Item Test</td>
<td>93.8</td>
<td>3.3</td>
<td>2.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Amsterdam Short Memory Test</td>
<td>99.3</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Automated Sequences Task</td>
<td>90.2</td>
<td>2.5</td>
<td>2.2</td>
<td>1.8</td>
<td>3.3</td>
</tr>
<tr>
<td>The b Test</td>
<td>92.0</td>
<td>4.0</td>
<td>3.3</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>CARB</td>
<td>95.3</td>
<td>2.2</td>
<td>1.8</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>CVLT/C Discriminability Index</td>
<td>37.0</td>
<td>5.8</td>
<td>20.7</td>
<td>20.7</td>
<td>15.9—37%</td>
</tr>
<tr>
<td>CVLT-II Effort Algorithm Wolf 2010</td>
<td>86.6</td>
<td>4.0</td>
<td>5.8</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>CVLT-II Forced Choice</td>
<td>27.2</td>
<td>7.2</td>
<td>29.0</td>
<td>21.4</td>
<td>15.2—37%</td>
</tr>
<tr>
<td>Dot Counting Test</td>
<td>87.0</td>
<td>3.6</td>
<td>6.5</td>
<td>2.9</td>
<td>0.0</td>
</tr>
<tr>
<td>MSVT</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-MSVT</td>
<td>85.1</td>
<td>5.1</td>
<td>3.3</td>
<td>5.4</td>
<td>1.1</td>
</tr>
<tr>
<td>Reliable Digit Span</td>
<td>34.8</td>
<td>8.0</td>
<td>13.8</td>
<td>22.1</td>
<td>21.4—44%</td>
</tr>
<tr>
<td>Rey-15 Item Test</td>
<td>66.3</td>
<td>17.0</td>
<td>9.1</td>
<td>6.2</td>
<td>1.4</td>
</tr>
<tr>
<td>TOMM</td>
<td>22.1</td>
<td>12.0</td>
<td>31.2</td>
<td>20.7</td>
<td>14.1—35%</td>
</tr>
<tr>
<td>Word Completion Memory Test</td>
<td>95.7</td>
<td>2.2</td>
<td>0.7</td>
<td>1.4</td>
<td>0.0</td>
</tr>
<tr>
<td>WMT</td>
<td>69.6</td>
<td>78.0</td>
<td>8.0</td>
<td>8.7</td>
<td>5.8</td>
</tr>
<tr>
<td>VSST</td>
<td>85.5</td>
<td>6.2</td>
<td>4.0</td>
<td>2.9</td>
<td>1.4</td>
</tr>
</tbody>
</table>

### Stand-Alone PVTs Investigated in Pediatric Populations

- **Pediatric PVT Reviews**
  
  

**Notes:** --- adequate evidence; — modest evidence; — no or conflicting evidence

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 (Lezak, 1983; Rey, 1941)
  - Martin, Haut, Stainbrook, & Franzen (1995); Rambo et al. (2015)
- 21-Item Test (Iverson, 1998)
- Computerized Assessment of Response Bias (Allen, Conder, Green & Cox, 1997)
  - Courtney, Dinkins, Allen, & Kuroski (2003); Harrison et al. (2014)
- Amsterdam Short-Term Memory Test (Schmand & Lindeboom, 2004)
  - Rienstra, Spaan, & Schmand (2010)
- Word Completion Memory Test (WCMT; Hilsabeck & LeCompte, 1997)
  - Rienstra, Spaan, & Schmand (2010)

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
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)

Boat

Water

Dog

Cat

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”
### WMT

Table 4. Summary of pediatric studies focused on the Word Memory Test

<table>
<thead>
<tr>
<th>Source</th>
<th>Population</th>
<th>N</th>
<th>Age Range</th>
<th>Mean Age (SD)</th>
<th>IR % Mean (SD)</th>
<th>DR % Mean (SD)</th>
<th>CNS %</th>
<th>% Passing *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rietveld et al. (2010)</td>
<td>Netherlands Community</td>
<td>48</td>
<td>7 – 12</td>
<td>9.9 (1.6)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>100%</td>
</tr>
<tr>
<td>Green et al. (2012)</td>
<td>Canada Clinical mixed &gt; 3rd grade reading level</td>
<td>360</td>
<td>--</td>
<td>13.4 (2.7)</td>
<td>95.9 (5.7)</td>
<td>95.9 (7.0)</td>
<td>93.3 (7.7)</td>
<td>90%</td>
</tr>
<tr>
<td>Courtney et al. (2003)</td>
<td>U.S. Clinical mixed – younger group</td>
<td>55</td>
<td>6 – 9</td>
<td>8.5 (1.2)</td>
<td>Average effort scores 74.2 (18.8)</td>
<td>Average effort scores 91.4 (10.4)</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Courtney et al. (2003)</td>
<td>U.S. Clinical mixed – older group</td>
<td>56</td>
<td>10 – 17</td>
<td>13.4 (2.0)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Leventhal &amp; Harrison (2012)</td>
<td>U.S. Clinical Learning Disability</td>
<td>63</td>
<td>11 – 14</td>
<td>12.2 (0.6)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>91%</td>
</tr>
<tr>
<td>Gunn et al. (2010)</td>
<td>Australia Simulation controls</td>
<td>50</td>
<td>6 – 11</td>
<td>8.7 (-1.13)</td>
<td>90.6 (7.6)</td>
<td>95.3 (6.1)</td>
<td>--</td>
<td>98%</td>
</tr>
</tbody>
</table>


### MSVT

Table 5. Summary of pediatric studies focused on the Medical Symptom Validity Test

<table>
<thead>
<tr>
<th>Source</th>
<th>Population</th>
<th>N</th>
<th>Age Range</th>
<th>Mean Age (SD)</th>
<th>IR % Mean (SD)</th>
<th>DR % Mean (SD)</th>
<th>CNS %</th>
<th>% Passing *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green et al. (2009)</td>
<td>Canada Community</td>
<td>56</td>
<td>7 – 11</td>
<td>9.2 (1.7)</td>
<td>98.6 (3.8)</td>
<td>98.6 (3.0)</td>
<td>97.6 (2.4)</td>
<td>96%</td>
</tr>
<tr>
<td>Green et al. (2009)</td>
<td>Brazil Community young</td>
<td>36</td>
<td>6 – 10</td>
<td>8.7 (1.4)</td>
<td>95 (5)</td>
<td>96 (4)</td>
<td>94 (8)</td>
<td>98%</td>
</tr>
<tr>
<td>Green et al. (2009)</td>
<td>Brazil Community old</td>
<td>34</td>
<td>11 – 15</td>
<td>12.4 (1.3)</td>
<td>96 (4)</td>
<td>100 (2)</td>
<td>96 (4)</td>
<td>98%</td>
</tr>
<tr>
<td>Green et al. (2012)</td>
<td>Canada Clinical mixed &gt; 3rd grade reading level</td>
<td>265</td>
<td>--</td>
<td>13.6 (2.3)</td>
<td>98.8 (3.7)</td>
<td>99.0 (4.3)</td>
<td>97.3 (5.8)</td>
<td>95%</td>
</tr>
<tr>
<td>Carone (2008)</td>
<td>U.S. Clinical mixed</td>
<td>38</td>
<td>--</td>
<td>11.8 (3.1)</td>
<td>98.6 (3.7)</td>
<td>97.8 (6.3)</td>
<td>96.7 (9.0)</td>
<td>95%</td>
</tr>
<tr>
<td>Kirkwood &amp; Kite (2010)</td>
<td>U.S. Clinical mild TBI</td>
<td>193</td>
<td>8 – 17</td>
<td>14.3 (2.3)</td>
<td>95.3 (5.3)</td>
<td>95.6 (5.4)</td>
<td>91.9 (4.8)</td>
<td>93%</td>
</tr>
<tr>
<td>Chabrol et al. (2007)</td>
<td>U.S. Social Security Disability applicants</td>
<td>25</td>
<td>6 – 16</td>
<td>11.2 (2.6)</td>
<td>86.4 (8.0)</td>
<td>84.2 (9.9)</td>
<td>87.8 (9.1)</td>
<td>93%</td>
</tr>
<tr>
<td>Blaskevitz et al. (2008)</td>
<td>Germany Simulation controls</td>
<td>51</td>
<td>6 – 11</td>
<td>8.9 (1.0)</td>
<td>98.6 (2.5)</td>
<td>99.6 (1.2)</td>
<td>98.2 (3.6)</td>
<td>98%</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)

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
Bottom line
- Probably appropriate in higher functioning children 11+ years
- Extreme caution in younger/lower functioning children
- Blaskewitz et al. (2008) and Cassie Green et al. (2014) suggest traditional cutoff scores quite insensitive to noncredible effort
- Results from Green et al. (2014) support adding Boone recognition trial
  - increased sensitivity considerably, without altering specificity, at least among the higher functioning 8-17 year olds

### 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
- Normed 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 Indicators

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

Embedded PVTs Investigated in Pediatric Populations

- Review

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

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

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

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

- Symptom Validity Scale
  - Chafetz et al. (2007; 2008)

- CNS Vital Signs
  - Brooks et al. (2014)

- ChAMP
  - Sherman & Brooks (2015)
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
    - < 5 has typically been associated with > 90% specificity, with sensitivity ranging from about 25% to 50%
  - Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994)
    - Calculated by summing the longest string of digits repeated without error over two trials under both forward and backward conditions
    - Cutoff of < 8 or < 7 has produced sensitivity values above 50% 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

The Value of the WISC-IV Digit Span Subtest in Detecting Noncredible Performance during Pediatric Neuropsychological Examinations†

Michael W. Kirkwood*, David D. Hargrave, John W. Kirk

Department of Physical Medicine & Rehabilitation, The Children’s Hospital, Colorado and University of Colorado School of Medicine, Aurora, CO, USA

*Corresponding author at: The Children’s Hospital, Connerison Program, 825, 1012 E. 16th Ave, Aurora, CO 80045, USA; Tel.: +720-775-1055; Fax: +720-775-1097.

Email address: kirkwood.michael@children.org (M.W. Kirkwood).

Accepted 21 April 2011

Abstract

In adult populations, research on methodologies to identify negative response bias has grown exponentially in the last two decades. Fewer studies have focused on methods appropriate for children. Although several recent studies have established the appropriateness of using standalone symptom validity tests with younger populations, a lack of pediatric work has generated embedded validity indicators. The present study examined the classification value of several scores derived from the WISC-IV Digit Span subtest. The sample consisted of 274 clinically referred mild traumatic brain injury patients aged 8 through 18 years. Fifteen percent of the participants failed both the Medical Symptom Validity Test and Test of Memory Malingering, which was used as the criterion for noncredible effort. For age-corrected scaled scores, a score of ≤ 5 resulted in the optimal cut-score, yielding sensitivity of 51% and specificity of 96%. For Reliable Digit Span, the optimal cut-score was ≤ 5, with sensitivity of 51% and specificity of 92%. Although only moderately sensitive, Digit Span scores are likely to have good utility in identifying noncredible performance in relatively high-functioning older children and adolescents. Indeed, classification statistics produced in this pediatric sample compare favorably with those produced in many real-world adult patients.

Keywords: Digit span; Reliable digit span; Wisc VI; Digit span; WISC IV; Symptom validity testing; Response bias; Postconcussion; Mild traumatic brain injury.
• 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; 6/7 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

Clinical Utility of Reliable Digit Span in Assessing Effort in Children and Adolescents with Epilepsy

Antoinette J. Welsh, H. Allison Bender, Lindsay A. Whittingham, Mursha Vasserman, William S. MacAllister.

- 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
BRIEF (and other domain-specific scales)

- No identified independent studies examining faking bad or negativity scales

<table>
<thead>
<tr>
<th>SVT</th>
<th>Never (%)</th>
<th>Rarely (%)</th>
<th>Sometimes (%)</th>
<th>Often (%)</th>
<th>Almost 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>29.0 → 48%</td>
</tr>
<tr>
<td>BRIEF Validity Indicators</td>
<td>27.2</td>
<td>7.2</td>
<td>12.0</td>
<td>21.0</td>
<td>32.6 → 54%</td>
</tr>
<tr>
<td>MMPI-A Indicators</td>
<td>47.1</td>
<td>15.6</td>
<td>13.4</td>
<td>13.4</td>
<td>10.5 → 24%</td>
</tr>
<tr>
<td>Personality Inventory for Youth Validity</td>
<td>86.6</td>
<td>4.3</td>
<td>3.9</td>
<td>3.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Indicators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trauma Symptom Checklist for Children</td>
<td>87.3</td>
<td>5.8</td>
<td>4.7</td>
<td>0.7</td>
<td>1.4</td>
</tr>
<tr>
<td>Validity Indicators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
N = 274; 8-17 year olds administered BASC2

Primary question
- What is the relationship between the BASC-2 validity indicators and PVT performance in a sample of real-world pediatric patients?

Hypothesis
- Children who failed MSVT would be more likely to have elevations on BASC-2 validity indices, most notably the F index

Self-Report BASC2 Validity Indicators
- F Index: designed to assess that a child responded in an inordinately negative fashion or was "faking bad"
- L index: designed to detect a response set that may be characterized as one of social desirability or "faking good"
- V index: consists of nonsensical items that may be marked because of carelessness or failure to cooperate or understand questions
- Response Pattern index: designed to identify forms that may be invalid because of distraction to item content (e.g., N-N-N-N; T-F-T-F-T-F)
- Consistency index: identifies cases where differing responses given to items usually answered similarly

<table>
<thead>
<tr>
<th>BASC-2 SRP Validity Scales</th>
<th>MSVT</th>
<th>Significant Test</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PASS</td>
<td>FAIL</td>
<td>Two-tailed, Fisher’s Exact Test</td>
</tr>
<tr>
<td>F Scale Within “Caution” or “Extreme Caution” range</td>
<td>4</td>
<td>3</td>
<td>.117</td>
</tr>
<tr>
<td>F Scale Within normal limits</td>
<td>220</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Response Pattern Within “Caution” or “Extreme Caution” range</td>
<td>1</td>
<td>1</td>
<td>.332</td>
</tr>
<tr>
<td>Response Pattern Within normal limits</td>
<td>223</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Consistency Scale Within “Caution” or “Extreme Caution” range</td>
<td>10</td>
<td>0</td>
<td>.217</td>
</tr>
<tr>
<td>Consistency Scale Within normal limits</td>
<td>214</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>L Scale Within “Caution” or “Extreme Caution” range</td>
<td>14</td>
<td>0</td>
<td>.081</td>
</tr>
<tr>
<td>L Scale Within normal limits</td>
<td>210</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>V Scale Within “Caution” or “Extreme Caution” range</td>
<td>2</td>
<td>0</td>
<td>.100</td>
</tr>
<tr>
<td>V Scale Within normal limits</td>
<td>222</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Any Validity Scale Within “Caution” or “Extreme Caution” range</td>
<td>29</td>
<td>4</td>
<td>.471</td>
</tr>
<tr>
<td>Any Validity Scale Within normal limits</td>
<td>195</td>
<td>46</td>
<td></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

<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>10.6</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>35.8</td>
<td>33.6</td>
<td>6.0</td>
</tr>
<tr>
<td>No firm conclusions can be drawn</td>
<td>9.3</td>
<td>16.2</td>
<td>37.4</td>
<td>31.3</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>40.8</td>
<td>35.1</td>
<td>5.7</td>
</tr>
<tr>
<td>Test results indicate inadequate engagement</td>
<td>15.9</td>
<td>15.9</td>
<td>40.5</td>
<td>23.1</td>
<td>4.5</td>
</tr>
<tr>
<td>Test results indicate poor compliance</td>
<td>18.9</td>
<td>20.5</td>
<td>37.9</td>
<td>18.2</td>
<td>4.5</td>
</tr>
<tr>
<td>Test results indicate exaggeration or feigning</td>
<td>28.3</td>
<td>31.7</td>
<td>32.8</td>
<td>4.9</td>
<td>2.3</td>
</tr>
<tr>
<td>Test results indicate malingering</td>
<td>64.9</td>
<td>28.7</td>
<td>5.7</td>
<td>0.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

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

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, Connery, & 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
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
   - Systemically

## 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>100</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</td>
<td>(mean: 11.8)</td>
<td>MSVT</td>
<td>5%</td>
</tr>
<tr>
<td>MacAllister, Nakhutina, 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></td>
<td>WMT</td>
<td>5%</td>
</tr>
<tr>
<td>Green et al. (2010)</td>
<td>Mixed Neuro/Dev</td>
<td>265</td>
<td></td>
<td>MSVT</td>
<td>3%</td>
</tr>
<tr>
<td>Kirk, Harris, Hutaff-Lee, Koelmay, Dinkins, &amp; Kirkwood (2011)</td>
<td>Mixed Neuro/Dev</td>
<td>100</td>
<td>5 – 16</td>
<td>TOMM</td>
<td>4%</td>
</tr>
<tr>
<td>Brooks (2012)</td>
<td>Mixed Neuro</td>
<td>100</td>
<td>6 – 19</td>
<td>VSVT</td>
<td>5%</td>
</tr>
<tr>
<td>Ploetz, Mosiewicz, Kirkwood, Sherman, &amp; Brooks (2014)</td>
<td>Mixed Neuro</td>
<td>266</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); Larson 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 Presentation</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% (26-30% PVT chance level or below)</td>
</tr>
</tbody>
</table>

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

- 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>Pass</th>
<th></th>
<th></th>
<th>Fail</th>
<th></th>
<th></th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>WASI</td>
<td>213</td>
<td>105.5</td>
<td>11.6</td>
<td>48</td>
<td>94.3</td>
<td>13.4</td>
<td>.106</td>
<td>.9</td>
</tr>
<tr>
<td>Vocabulary T score</td>
<td>212</td>
<td>33.6</td>
<td>10.8</td>
<td>48</td>
<td>20.7</td>
<td>10.9</td>
<td>.008</td>
<td>.3</td>
</tr>
<tr>
<td>Matrix Reasoning T score</td>
<td>215</td>
<td>52.1</td>
<td>7.3</td>
<td>50</td>
<td>41.0</td>
<td>10.6</td>
<td>.007</td>
<td>1.4</td>
</tr>
<tr>
<td>CVLT-C</td>
<td>186</td>
<td>53.0</td>
<td>8.1</td>
<td>40</td>
<td>46.6</td>
<td>11.4</td>
<td>.008</td>
<td>0.7</td>
</tr>
<tr>
<td>Total Learning Trials 1-5 T score</td>
<td>188</td>
<td>0.34</td>
<td>0.8</td>
<td>40</td>
<td>-0.48</td>
<td>1.3</td>
<td>.040</td>
<td>0.9</td>
</tr>
<tr>
<td>Recognition Discriminability T score</td>
<td>186</td>
<td>0.18</td>
<td>0.6</td>
<td>40</td>
<td>-1.29</td>
<td>1.8</td>
<td>.002</td>
<td>1.6</td>
</tr>
<tr>
<td>WISC-IV</td>
<td>224</td>
<td>9.9</td>
<td>2.9</td>
<td>51</td>
<td>6.4</td>
<td>3.2</td>
<td>.001</td>
<td>1.2</td>
</tr>
<tr>
<td>Digit Span scaled score</td>
<td>207</td>
<td>9.7</td>
<td>5.3</td>
<td>45</td>
<td>6.4</td>
<td>3.1</td>
<td>.06</td>
<td>0.6</td>
</tr>
<tr>
<td>Coding scaled score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Programmed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominate hand z score</td>
<td>213</td>
<td>-0.23</td>
<td>1.4</td>
<td>45</td>
<td>-1.7</td>
<td>2.2</td>
<td>.026</td>
<td>0.9</td>
</tr>
<tr>
<td>Non-dominant hand z score</td>
<td>215</td>
<td>-0.43</td>
<td>1.5</td>
<td>45</td>
<td>-1.6</td>
<td>2.2</td>
<td>.003</td>
<td>0.7</td>
</tr>
<tr>
<td>Woodcock-Johnson III</td>
<td>191</td>
<td>100.2</td>
<td>9.7</td>
<td>45</td>
<td>97.0</td>
<td>22.0</td>
<td>.347</td>
<td>0.3</td>
</tr>
<tr>
<td>List-Word Identification scaled score</td>
<td>216</td>
<td>5.6</td>
<td>6.1</td>
<td>50</td>
<td>11.4</td>
<td>10.9</td>
<td>.006</td>
<td>0.8</td>
</tr>
<tr>
<td>Arithmetic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orientation 1 to 20</td>
<td>172</td>
<td>4.7</td>
<td>1.4</td>
<td>44</td>
<td>9.6</td>
<td>12.5</td>
<td>.043</td>
<td>0.9</td>
</tr>
<tr>
<td>Days of week</td>
<td>250</td>
<td>2.3</td>
<td>1.2</td>
<td>47</td>
<td>5.4</td>
<td>5.1</td>
<td>.000</td>
<td>1.2</td>
</tr>
<tr>
<td>Months of year</td>
<td>214</td>
<td>8.1</td>
<td>4.4</td>
<td>47</td>
<td>12.0</td>
<td>6.8</td>
<td>.000</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Implications of PVT Failure for Interpreting Symptom Report

- Several studies with adults have also suggested that performance on PVTs has significant effect on postconcussive symptom report after mild TBI
  - Lange et al. (2010)
  - Iverson et al. (2010)
  - Tsanadis 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, Connery, Baker, & Grubenhoff (2014)
• 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