Performance and Symptom Validity

Presentation to the Institute of Medicine Committee on Psychological Testing Including Symptom Validity Testing

Glenn J. Larrabee, Ph.D. ABPP
Psychological and Neuropsychological Testing

• Requires accurate self report on tests such as the MMPI-2

• Requires a valid performance on tests of ability, such as the Wechsler Intelligence and Memory scales

• Since both self report and performance can be controlled by the examinee, it is critical that symptom validity and performance validity is evaluated in each assessment
Performance Validity Tests or PVTs

- PVTs tell you whether the examinee is providing an accurate measure of their actual abilities.
- These can be stand-alone like the Test of Memory Malingering (TOMM) or Word Memory Test (WMT).
- These can be embedded/derived measures based on neurologically atypical patterns or levels of performance on standard tests like the Auditory Verbal Learning Test, Finger Tapping, or Reliable Digit Span.
Symptom Validity Tests or SVTs

• SVTs tell you whether the examinee is providing an accurate report of their actual symptom experience
• SVTs are included on personality tests like the MMPI-2-RF (F-r, Fp-r, Fs, FBS-r, RBS)
• SVTs have also been developed for pain scales such as the Modified Somatic Perception Questionnaire or Pain Disability Index
PVT Failure Distorts Expected Relationships with Validity Criteria

• GPA does not show the expected relationship with IQ until those failing PVTs are excluded (Greiffenstein & Baker, 2003)
• Olfactory identification does not correlate with TBI severity until those failing PVTs are excluded (Green et al., 2003)
• Memory complaints only correlate with memory performance in those failing PVTs (Gervais et al., 2008)
• Memory test scores correlate .49 with hippocampal volume in MCIs who pass PVTs, but correlate -.11 for those who fail PVTs (Rienstra et al., 2013)
PVT Failure Distorts Expected Relationships with Validity Criteria

- Memory performance does not differ between examinees with and without CT scan abnormalities until those failing PVTs are excluded (Green, 2007)
- A TBI/neurologically impaired group did not differ on neuropsychological tests from psychiatric, pain or mTBI subjects until those failing PVTs were excluded (Green et al., 2001)
- Neuropsychological performance was associated with presence/absence of brain injury only in those Ss passing PVTs (Fox, 2011)
Frequency of PVT failure in Settings with External Incentive

• 40% in litigation/compensation-seeking mTBI (Larrabee, 2003)
• 54.3% in criminal defendants (Ardolf et al. 2007)
• 48.5% in Social Security claimants (Chafetz, 2008)
• 40% in persons claiming environmental or toxic exposure (Greve et al. 2006)
## Costs for SSD (SSI and SSDI) in Billions
Chafetz & Underhill, 2013

<table>
<thead>
<tr>
<th>Frequency of PVT Failure</th>
<th>Mental Disorders</th>
<th>Musculoskeletal Disorder (Chronic Pain)</th>
<th>Combined Mental Disorders/Musculoskeletal Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>40%</td>
<td>$20.022 Billion</td>
<td>$14.176 Billion</td>
<td>$34.198 Billion</td>
</tr>
<tr>
<td>50%</td>
<td>$25.028 Billion</td>
<td>$17.72 Billion</td>
<td>$42.748 Billion</td>
</tr>
</tbody>
</table>
Research on SVTs and PVTs In Psychology and Neuropsychology Has Lengthy and Extensive History

- F scale included in MMPI (1943)
- Rey (1941) Dot Counting Test
- Rey (1964) 15-item Test
- Research on feigned BVRT performance (Benton & Spreen, 1961)
- Two alternative forced choice testing (Pankratz, Fausti & Peed, 1975)
- Pattern analysis on standard neuropsychological tests (Heaton, Smith, Lehman & Vogt, 1978)
- Slick et al. (1999) diagnostic criteria for malingered neurocognitive dysfunction
- Bianchini et al. (2005) criteria for malingered pain related disability
Explosion in PVT and SVT Research 1990 to Present (Sweet & Guidotti Breting, 2013)


• 14 Meta-analytic reviews of PVTs and SVTs, 1991-2011

Position Papers from Major Organizations

• NAN Position Paper, 2005: Symptom exaggeration or fabrication occurs in a sizeable minority of exams, with greater prevalence in forensic contexts; use of PVTs/SVTs maximizes confidence in results and is medically necessary

• AACN Consensus Statement, 2009: 30 experts; PVT/SVT assessment is important and necessary, particularly in secondary gain contexts, but also in routine clinical practice
How are PVTs created?
Simulation Design

• Simulation design usually compares two groups of Ss on the PVT being developed:
  1) a group of non-injured persons asked to believably feign deficits in an imagined personal injury suit
  2) Persons with moderate/severe TBI who do not have external incentive
How are PVTs Created?

Known Groups or Criterion Groups Design

- Usually compares two groups on the PVT being studied:

1) a litigating/compensation-seeking group of uncomplicated mTBI (normal CT scan, no or brief LOC, limited PTA), who also fail 2 or more PVTs independent of the PVT being studied

2) a group of Ss with moderate/severe TBI who do not have any external incentive
PVTs, SVTs and Effect of Minimizing the False Positive Rate

• Use of non-litigating patients with moderate and severe TBI, with history of coma and/or CT scan abnormalities, plus a PVT/SVT cutoff with ≤10% false positive (FP) rate
• Specifying the characteristics of Ss who are false positive
• Because of the attempt to minimize false positive error, sensitivity is notably lower
• PVT meta-analysis of Vickery et al. (2001) found mean sensitivity of .56 with mean specificity of .95
• Sollman and Berry (2011) found mean sensitivity of .69 with mean specificity of .90
PVTs are Easy to Pass for Examinees who have Bona Fide Problems

- Mean TOMM correct is 98.7% for aphasics
- Mean TOMM correct is 98.3% for severe TBIs (>1 day up to 3 months of coma); 1 S with GSW, 38 days coma, & right frontal lobectomy scored perfectly (Tombaugh, 1996)
- 3 patients with bilateral hippocampal damage due to hypoxia passed the PVT trials of the Word Memory Test (Goodrich-Hunsaker & Hopkins, 2009)
- 94% of Ss with temporal lobe damage (57% left/43% right), intractable epilepsy, & AVLT delay = T39, pass the Effort Index of the AVLT (Silverberg & Barrash, 2005)
- Performance on 3 PVTs and 1 SVT was not related to presence or absence of CT/MRI lesions or to frontal vs. non-frontal location (McBride et al., 2013)
PVTs are Easy to Pass in Patients with Bona Fide Problems

- Cold pressor does not impact performance on Reliable Digit Span or on the TOMM (Etherton, Bianchini, Ciota et al., 2005; Etherton, Bianchini, Greve et al., 2005)
- “Diagnosis Threat” does not affect WMT performance (Suhr & Gunstad, 2005)
- Depression (Rees et al., 2001), depression and anxiety (Ashendorf et al., 2004) and depression with chronic pain (Iverson et al., 2007) do not affect TOMM performance
- 100% of fibromyalgia and rheumatoid arthritis Ss not seeking compensation passed the CARB (Gervais, Russell et al., 2001)
When it comes to effort tests, easy means... *easy*!

Imaging at age 1.
- Large head
- Right hemiparesis
- Seizures beginning age 5.
- Abnormal EEGs; lots of meds

Testing age 9.5; completing third grade.

- Full Scale IQ: 69
- Word Reading: 21st percentile
- Reading comprehension: 1st percentile
- Math computation: 1st percentile
- Many other tests; most low scores.

When it comes to effort tests, easy means...easy!

“Easy” subtests of Medical Symptom Validity Test

IR 100%
DR 100%
Cons 100%

“Easy” subtests of Word Memory Test

IR 95%
DR 100%
Cons 95%

The Positive Likelihood Ratio
Grimes & Schulz 2005

• LR+ is defined as sensitivity/1 – specificity
• LR+ gives likelihood the score came from the group with the condition of interest (COI) as opposed to the group without the COI
• LR+ multiplied by the base rate odds gives the post-test odds, which can be converted back to a diagnostic probability by the formula odds/odds + 1
• Based on the Vickery et al. PVT meta analysis, LR+ is .56/.05 or 11.2
• Based on the Sollman & Berry PVT meta analysis, LR+ is .69/.10 or 6.9
LR+ for Various Clinical Disorders: Heaton et al., 2004; Center For Evidence Based Medicine, Toronto, ktclearinghouse.ca/cebm/glossary/lr 6/03/14

- Brain damage vs. no brain damage using AIR from Halstead Reitan (Heaton et al. 2004), LR+ = .771/.146 = 5.28
- Alzheimer’s positive APOE, LR+ = 2.0
- Myocardial Infarction, Cardiac Specific Troponin T, LR+ at 0-2 hours = 6.3
- Chronic Obstructive Airway Disease by spirometry, LR+ = 7.3
Improving Diagnostic Accuracy by Using Multiple Indicators (Larrabee, 2008)

- LR+ can be chained, if the individual PVTs and SVTs are independent and either weakly correlated or uncorrelated.
- Consider 2 independent, uncorrelated PVTs, each with a sensitivity of .50 and specificity of .90, and a base rate probability of PVT failure of .40 (yielding base rate odds of .40/1-.40 or .67).
- Post test odds after failing the 1st PVT are .67 x 5.0 = 3.35, for a post-test probability of 3.35/4.35 = .77.
- The post-test odds of 3.35 can now be used to premultiply the LR+ for a 2nd failed PVT, which also has an LR+ of 5.0, yielding new post-test odds of 16.75, for a post-test probability of 16.75/17.75 or .94.
Use of Multiple PVTs Does Not Negatively Impact the Per Test False Positive Rate

• Berthelson et al. (2013) used Monte Carlo simulation to estimate failure rates for subsets of PVTs failed at either a 10% or 15% per-test false positive rate

• Berthelson et al. found what they interpreted as excessive false positive rates associated with use of multiple PVTs
Actual PVT Failure Rates in Clinical Patients Contradicts Berthelson et al.

• Davis and Millis (2014) found that increasing the number of PVTs administered did not show a significant association with PVT failure.

• Davis and Millis attributed the difference in their findings compared to those of Berthelson et al. to violation of the Monte Carlo assumption of multivariate normality.
Why Monte Carlo Simulated Data do not Match Actual PVT False Positive Rates

- Berthelson et al. Monte Carlo simulated data creates variables with a mean of 0 and SD of 1
- These create simulated data that follow the standard normal curve
- PVT and SVT data do not follow the normal curve, as they are skewed, with performance commonly at ceiling
Examples of Skewed Data From the TOMM Manual

• 21 aphasic subjects averaged 98.7% correct on Trial 2, with 16 achieving perfect scores
• 22 severe TBI patients (> 1 day to 3 months of coma) averaged 98.2% correct on Trial 2 with 14 achieving perfect scores
• These data demonstrate skewed distributions with performances occurring at ceiling, a common finding with PVTs
Percentage Failing ≥2 and ≥3 PVTs at a 10% per PVT FP rate for 5 and 7 PVTs Larrabee 2014

<table>
<thead>
<tr>
<th>Study</th>
<th>≥2/5</th>
<th>≥3/5</th>
<th>≥2/7</th>
<th>≥3/7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berthelson Monte Carlo</td>
<td>11.5</td>
<td>3.6</td>
<td>17.5</td>
<td>7.3</td>
</tr>
<tr>
<td>ACS TBI, neuro, psych, &amp; developmental N = 371</td>
<td>6.0</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larrabee, TBI, neuro, psych N = 54</td>
<td>5.6</td>
<td>0.0</td>
<td>11.1</td>
<td>3.7</td>
</tr>
<tr>
<td>Schroeder non-psychotic N = 178</td>
<td></td>
<td></td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Schroeder psychotic N = 104</td>
<td></td>
<td></td>
<td>7.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Putting It All Together

• Failure of multiple PVTs and SVTs indicates that there is a high probability of invalid data
• By itself, this indicates that low scores in an evaluation are more likely due to invalid performance, particularly with absence of indicators for severe impairment such as history of coma
• Similarly, normal range scores themselves may be low estimates of actual level of ability
Putting it All Together

• PVT and SVT failure alone does not equate to malingering

• Multiple PVT and SVT failure, in the context of external incentive, with no clear evidence of neurologic, psychiatric or developmental contributions to test performance, meets general criteria for probable malingering (Slick et al., 1999; Bianchini et al. 2005)

• Below chance on 2-alternative forced choice testing, in the context of external incentive, meets criteria for definite malingering, Pankratz’ (1990) “smoking gun of intent”
Putting it All Together
Larrabee et al 2007 Chapter 13

• Definite malingerers (< chance) perform similarly to non-injured simulators, establishing < chance as intentional

• Definite malingerers (< chance) perform similarly to criterion group probable malingerers (multiple PVT/SVT failures), establishing validity of probable malingering criteria as supporting intentionally poor performance

• Dose effect (Bianchini et al., 2006) showing positive correlation of malingering with increasing external incentive shows that PVT/SVT failure is reinforced by pursuit of external incentives
## Percentage of Clinical and Malingering Cases Failing PVTs and SVTs Larrabee 2014

<table>
<thead>
<tr>
<th># PVT &amp; SVT Failures</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Cases</td>
<td>51.9%</td>
<td>37.0%</td>
<td>7.4%</td>
<td>3.7%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>n=54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malingering Cases</td>
<td>0</td>
<td>2.4%</td>
<td>9.8%</td>
<td>24.4%</td>
<td>26.8%</td>
<td>19.5%</td>
<td>17.1%</td>
<td>0</td>
</tr>
<tr>
<td>n=41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Sensitivity and Specificity for Larrabee 2014 Multiple Indicator Paper

<table>
<thead>
<tr>
<th>PVT/SVT Failure</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Likelihood Ratio +</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2 of 7</td>
<td>97.6</td>
<td>88.9</td>
<td>8.79</td>
</tr>
<tr>
<td>≥3 of 7</td>
<td>87.8</td>
<td>96.3</td>
<td>23.73</td>
</tr>
<tr>
<td>≥4 of 7</td>
<td>63.4</td>
<td>100.0</td>
<td>Cannot compute LR+, however PPP = TP/TP + FP, = 100%</td>
</tr>
</tbody>
</table>
Improving Diagnostic Accuracy Results from Controlling per-test False Positive rate and use of Multiple PVTs

- PVT and SVT research focuses on keeping the False+ (FP) rate at 10% or less *per test* in subjects with significant neurologic, psychiatric and developmental problems
- Current PVT and SVT research specifies the characteristics of FP cases in derivation studies
- FP typically occur in Ss who have undeniably severe deficits, often requiring 24 hour supervised care
- Requirement of multiple PVTs and SVTs improves diagnostic accuracy by reducing False Positive and False Negative rates
Summary on PVTs and SVTs

• The science behind PVTs and SVTs dates back over 70 years, with extensive research in the past 20 years on both stand alone and embedded/derived measures.

• Diagnostic accuracy has focused on keeping the pre-test false positive rate (FP) at 10% or less which is further enhanced by requiring use of multiple independent PVTs/SVTs.

• PVT/SVT failure is rare in non-demented patients; rare in well-motivated patients, but frequent (40-50% of the time) in patients having external incentives.

• Costs for mental disorder and pain disability based on invalid data are in the billions of dollars per year.
References


References


References


References


References


References


References


References
