Innovations in Performance and Symptom Validity Assessment

Douglas Whiteside, PhD, ABPP/CN
• University of Minnesota
Michael Basso, PhD, ABPP/CN
• Mayo Clinic-Rochester
Ryan Schroeder, PsyD, ABPP/CN
• Wichita VA Medical Center
Jason Soble, PhD, ABPP/CN
• University of Illinois, Chicago
Patrick Armistead Jehle, PhD, ABPP/CN
• Munson Army Medical Center

AACN 2025

1

Financial Disclosure

We have no financial relationships to disclose

AACN 2025

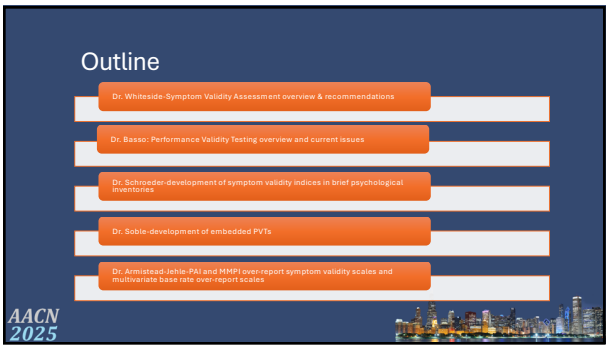
2

Learning Objectives

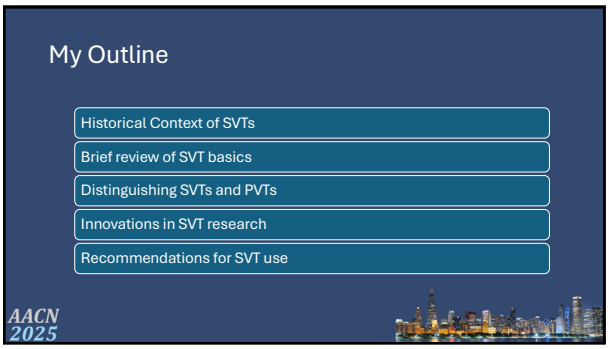
1. Participants will be able to describe empirical paradigms for developing and cross-validating validity tests among diverse clinical examinees.
2. Participants be able to identify validity test cutoff scores appropriate for use with various individuals including those with cognitive impairment and those of diverse ethnoracial populations.
3. Participants will be able to elaborate novel methods of measuring symptom and performance validity and illustrate their implications for practice and future innovation.

AACN 2025

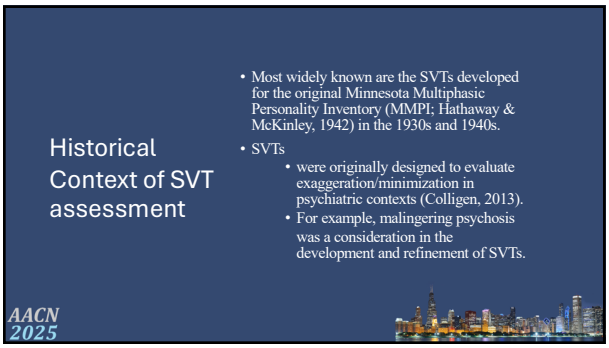
3


Outline
Dr. Whiteside-Symptom Validity Assessment overview & recommendations
Dr. Basso: Performance Validity Testing overview and current issues
Dr. Schroeder-development of symptom validity indices in brief psychological inventories
Dr. Soble-development of embedded PVTs
Dr. Armistead-Jehle-PAI and MMPI over-report symptom validity scales and multivariate base rate over-report scales
AACN 2025

4


My Outline
Historical Context of SVTs
Brief review of SVT basics
Distinguishing SVTs and PVTs
Innovations in SVT research
Recommendations for SVT use
AACN 2025

5


Historical Context of SVT assessment
Most widely known are the SVTs developed for the original Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1942) in the 1930s and 1940s.
SVTs
were originally designed to evaluate exaggeration/minimization in psychiatric contexts (Colligen, 2013).
For example, malingering psychosis was a consideration in the development and refinement of SVTs.
AACN 2025

6

Other personality measures use SVTs

- Personality Assessment Inventory (PAI; Morey, 2007), including newly developed SVTs like the Cognitive Bias Scale (CBS; Gaasedelen et al., 2019)
- Millon family of scales such as the Millon Clinical Multiaxial Inventory-IV (MCMI-IV; Millon et al., 2019)
- Successors to the MMPI, (i.e., the MMPI-2-RF and MMPI-3), implemented new SVTs (RBS, FBS).
- Most SVTs have been developed within broad based personality measures rather than as standalone measures of response bias.

AACN 2025

7

Brief review- types of SVTs

- Non-Content Responding
  - *nonresponding*, when test takers simply do not answer items.
  - *random responding*, when test takers respond randomly to items
  - *fixed responding*, when the test takers respond in the same way regardless of item content.

AACN 2025

8

Examples of non-content SVTs

- Non-content SVTs:
  - MMPI include TRIN and VRIN
  - PAI include ICN and INF

AACN 2025

9

Content SVTs

overreporting measures	assess if test takers are endorsing more distress or dysfunction than they are actually experiencing
under-reporting measures	assesses if test takers are minimizing or denying psychological issues

AACN 2025

10

Examples of content SVTs

Under-reporting scales:	• MMPI: K and L scales • PAI-PIM
Over-reporting scales:	• MMPI: F family of scales • PAI: NIM, MAL

AACN 2025

11

New types of SVTs

- Focus on exaggeration or embellishment of cognitive and physical symptoms, rather than traditional psychiatric symptoms
  - Examples
    - MMPI-2/2-RF: FBS/FBS-r for the MMPI-2/2-RF, Ben-Porath & Tellegen, 2008;
    - PAI: Cognitive Bias Scale (CBS; Gaasedelen et al., 2019),
  - These are more typically related to neuropsychological evaluations

AACN 2025

12

SVTs and PVTs

- Larrabee (2012) articulated PVT-SVT distinction
  - PVTs assess the credibility of task performance on objective cognitive measures
  - SVTs assess the accuracy of symptom complaints,
    - a subjective focus to the measures.
- Prior to that, it was common to refer to cognitively based measure of performance validity as SVTs.
- This has been an effective conceptualization

AACN 2025

13

SVT-PVT Relationship

- Using multiple regression and the PAI, Boress et al. (2024) found that the relationship is complex
  - varies in different clinical samples.
  - **Overall Sample**: CBS but not NIM or MAL predicted PVT (TOMM) performance
  - **Subsamples**:
    - mTBI-same result-CBS but not NIM/MAL
    - ADHD-no relationship between any PAI SVT and PVT performance
    - Psychiatry-all SVTs were related to PVT performance (CBS large effect size, NIM and MAL medium effect size)
  - Thus, PVTs and SVTs may be thought of as siblings: clearly related but also distinct individuals with complex inter-relationships.

AACN 2025

14

Sample	Outcome Measures	Model	R^2	R^2 change	F Change	Sig F-change
Full (*N* = 553)	TOMM	CBS	0.107	0.108	67.004	<0.001
		CBS, NIM, MAL	0.117	0.010	4.423	0.052
mTBI (*N* =79)	TOMM	CBS	0.059	0.059	5.654	0.020
		CBS, NIM, MAL	0.119	0.050	2.981	0.056
ADHD/LD (*N* = 92)	TOMM	CBS	0.002	0.002	0.146	0.703
		CBS, NIM, MAL	0.014	0.012	0.548	0.580
Psychiatric 1 (*N* = 382)	TOMM	CBS	0.141	0.141	60.733	<0.001
		CBS, NIM, MAL	0.184	0.043	9.788	<0.001
Psychiatric 2 (*N* = 382)	TOMM	NIM, MAL	0.160	0.160	35.273	<0.001
		NIM, MAL, CBS	0.184	0.024	10.776	<0.001

AACN 2025

15

SVT-PVT distinction

- Basso et al:
  Used SEM to evaluate relationship between PVTs and SVTs with inpatient mood disorder patients
  - Best model fit indicated that SVTs and PVTs were separate constructs
- Other research suggests SVTs can predict PVT performance
- However, SVTs should NOT be considered proxies for PVTs. There is enough distinctness in constructs that both need to be used.

AACN 2025

16

Examples of innovative SVT research-specific populations

- PTSD:
- Schroeder et al.-PTSD Checklist for the Diagnostic and Statistical Manual-5 (DSM-5; PCL-5) validity indices in a PTSD sample and found initial support for the SVTs in the PCL-5.
- ADHD:
- Finley et al. (2024) cross-validity study of Clinical Assessment of Attention Deficit-Adult (CAT-A).
  - examined Negative Impression (NI) and Positive Impression (PI) scale

AACN 2025

17

Examples of innovative SVT research-specific populations

- Criminal settings-Denney Competency Related Test (D-CRT)
- Denney et al. (2024) developed a new stand-alone measure of malingering specifically for criminal population.
  - Found their new measure has promising psychometric properties
  - Test-retest was good (.83)
  - Internal consistency excellent (KR-20>.91)
  - Convergent validity good (MacArthur)
  - Classification accuracy (ROC) very good (AUC=.945)

AACN 2025

18

Recommendations for the use of SVTs

- SVTs should be employed when complaints of emotional, cognitive, or physical symptoms are salient.
  - particularly recommended when psychiatric issues are prominent in the referral question
  - medico-legal cases where examinees may embellish or exaggerate reports of their physical malfunctioning

AACN 2025

19

Recommendations for the use of SVTs

- Use of SVTs should be given consideration during the assessment of younger patients who present with vague cognitive, physical, or emotional complaints.
  - Dementias are less common in younger patients than in individuals who have reached the seventh decade of life
  - the base rate of neurologically based cognitive dysfunction may be low, and a thorough differential diagnostic assessment requires careful evaluation of emotional functioning.

AACN 2025

20

Recommendations for the use of SVTs

- Not all patients referred for neuropsychological evaluations present with an extensive psychiatric history, so some may not need as extensive a psychological evaluation as others.
  - dementia and with no known external incentive
  - Some of these patients may not be able to complete broad band measures that include SVTs

AACN 2025

21

Recommendations for the use of SVTs

- Regardless, neuropsychologists are advised to carefully consider contextual, cultural, and demographic variables when deciding whether to include measures with SVTs.

AACN 2025

22

Recommendations for the use of SVTs

- As with PVTs, the use of SVTs is warranted whenever external incentives are known or suspected, regardless of the type of setting (clinical versus medico-legal).
  - Using SVTs even when the patient denies external incentive is likely to be helpful since patients are not always honest about such issues.

AACN 2025

23

Recommendations for the use of SVTs

- For any evaluation in a medico-legal setting, SVTs are strongly recommended.
  - The combined use of SVTs and PVTs is indicated since there is considerable unique variance accounted for by each type of measure, and this which varies by populations

AACN 2025

24

Conclusions

- SVTs are a critical component of neuropsychological evaluations
- Long history of SVT use in psychological/neuropsychological evaluations
- Need more innovative research on SVTs
  - Brief measures with good sensitivity and appropriate specificity
  - This is particularly critical for issues with a high base rate of external incentives
    - ADHD
    - Criminal forensics
    - PTSD

AACN 2025

25

MAYO CLINIC

AMERICAN ACADEMY OF CLINICAL NEUROPSYCHOLOGY CONFERENCE 2025

Performance Validity Testing: Considerations and Challenges

Michael R. Basso, Ph.D., L.P.
Professor and Consultant
Chicago, Illinois
June 14, 2025

1

Conflicts of Interest

- None

2

Objectives

- Review the history of performance validity assessment in clinical neuropsychology
- Highlight salient concerns regarding the current state of performance validity assessment
- Offer considerations of what should be addressed

3

Overview

- How did we get here?
- Where are we now?
  - Accomplishments and concerns
- Where should we go?
  - Recommendations for practice and research

4

How Did We Get Here?

- Before the 1990s
  - Performance validity rarely assessed (cf. Lezak, 1979)
  - Malingering was presumed unlikely in clinical evaluations (Bilder, 1986)
  - Prior to the 1990s, less than 10% of publications in neuropsychology journals concerned performance validity assessment (Martin et al., 2015)

5

Where Are We Now?

- Assessment of performance validity has become *de rigueur*
  - Practice guidelines render assessment the norm rather than the exception (e.g., Heilbronner et al., 2009; Sweet et al., 2021)
  - 20% of publications in neuropsychology journals concern performance validity (Martin et al., 2015)
  - Even non-psychologists are aware of performance validity measures
    - Lawyers acknowledge coaching clients to avoid detection (Lippa, 2017)
  - From rudimentary measures such as the Hiscock and Hiscock forced choice measure to multiple:
    - Standalone indices: TOMM, WMT, WCT, DCT
    - Embedded indices: ACS performance validity matrix, logical memory rarely missed index, CVLT FCI
    - Some of these even have population specific norms
  - Arguably, PVTs have exerted the biggest change upon clinical practice in the past thirty years

6

We've come a long way, but…

7

Where Should The Journey Begin?

- Existing PVTs are now recognized outside of our guild
  - New methods are necessary to avoid compromise or obsolescence
    - Patrick et al. (2024)
      - Pupillometry as a non-cognitive indicator of simulation
      - Difficult to elude detection
    - Basso et al. (2024)
      - Perceptual memory as a performance validity indictor
      - It's difficult to forget implicit memories

8

Where Should The Journey Begin?

- Population specific norms—Does one size fit all?
  - Boone et al. (2002)
    - Dot Counting Test includes norms for seven clinical populations
      - Different patients have markedly different cutoffs
  - Advanced Clinical Solutions (2008)
    - Specific norms for ten different clinical populations, and they vary considerably
  - Corriveau-Lecavalier et al. (2022)
    - Examined PVT accuracy in 17 patients with dysexecutive form of AD
      - Precocious onset (40s and 50s), impaired cognition but relatively preserved ADLs
      - Biomarker confirmed disease
      - Half of the patients failed the TOMM, and 25% had been diagnosed as malingering by clinicians during previous examinations
- To avoid harm to patients, specificity must be emphasized.
  - This probably requires population specific norms

9

Where Should The Journey Begin?

- Changing population demographics may degrade utility of existing norms
  - Denning & Horner (2024)
    - Compared 473 White and 58 African-American veterans on TOMM, MSVT and five embedded PVTs
    - After accounting for age, education, and sex:
      - PVT false positives were higher in African-Americans
      - Especially pronounced differences on embedded timed tasks (e.g., TMT-A)
- This raises concerns about the impact of other demographic characteristics
- More diverse norms should be obtained

10

Where Should The Journey Begin? A Fundamental Consideration

- What do PVTs measure?
  - Malingering? (Tombaugh, 1996)
    - Malingering requires external incentive
    - But…PVTs may be failed in the absence of an external incentive
    - Malingering is a potentially inaccurate or harmful diagnosis in clinical settings
  - "…engagement, symptom magnification, effort, or test validity…" (AACN-Guilmette et al. (2020)
    - Effort and engagement?
      - Some examinees might try pretty hard to simulate or exaggerate cognitive dysfunction
    - Symptom magnification?
      - That's more of a SVT instead of PVT function
    - PVTs measure performance validity (i.e., test validity)
      - Consider: " . . . intelligence as a measurable capacity must at the start be defined as the capacity to do well in an intelligence test . . . . measurable intelligence is simply what the tests of intelligence test (emphasis added) . . . " Boring (1923)
      - This is a tautological and unscientific definition
- The existing definitions are unsatisfactory

11

What Is A Defensible Definition?

- The empirical approach
  - Consider SVTs from the MMPI and PAI
    - Infrequently endorsed items reflect a pattern of improbable responding
    - High rates of improbable responses implies bias/error/invalidity
  - PVTs establish cutoffs upon base rate responses of patients/examinees
    - Too many improbably bad performances imply a pattern of biased responding
- A suggested operational definition: PVTs measure improbable cognitive responses
  - Multiple PVT failures reveal a pattern of improbable responding during the examination, raising doubts about the meaningfulness of performance on neuropsychological tests
- This definition offers a clear, discrete, and defensible operational definition of what is measured by PVTs

12

A Corollary To Consider

- If PVTs and SVTs measure improbable responding, can we equate them as a generic form of bias?
  - Is one form of improbable responding the same as another?

13

PVTs and SVTs: One or Two Constructs?

- Examinees who fail PVTs are more likely to fail SVTs and vice versa (Boress et al., 2024; Tombaugh, 1996; Whitney et al., 2008)
  - If a PVT is failed, then you don't need to administer a SVT
- Alternatively, there are indications that PVTs and SVTs measure different sources of variance (Ord et al., 2021; Van Dyke et al., 2013)
  - Failure of a PVT does not signify inevitable SVT failure

14

A Study to Determine Homogeneity or Heterogeneity

- 82 psychiatric inpatients diagnosed with unipolar, bipolar, or schizoaffective depression
- Administered WMT and MMPI-2
- WMT and MMPI validity scales entered into CFA
  - 2 models compared
    - Free estimation of variance between PVT and SVT constructs
    - Fixed relationship between PVT and SVT construct as 1.0 correlation

15

A Study to Determine Homogeneity or Heterogeneity

Table 1. Participant demographics.

Variable	Total Sample (N = 82)	Min.	Max.
Sociodemographic			
Age (years)	29.2 (12.1)	20	63
Gender (% female)	67.1		
Education (years)	12.3 (2.3)	8	19
Race (%)			
White	76.8		
Black	8.5		
Hispanic	1.2		
American Indian	7.3		
Other	6.1		
Primary Diagnoses (%)			
MDD	62.2		
MDD Psychotic	11.0		
Bipolar Depressed	12.2		
Bipolar Depressed Psychotic	2.4		
Schizoaffective Depressed	12.2		
Validity Performance			
WMT IR	88.8 (10.7)	55.0	100.0
WMT DR	89.7 (11.2)	42.5	100.0
WMT CNS	86.4 (11.5)	52.5	100.0
RBS	13.2 (4.1)	3	21
HHI	10.1 (3.1)	2	15
FBS	24.5 (5.4)	12	36
Failure Rates			
Invalid WMT	39.0%		
Invalid RBS	24.4%		
Invalid FBS	32.9%		
Invalid HHI	34.1%		

MDD = Major Depressive Disorder, WMT = Word Memory Test, RBS = Response Bias Scale, HHI = henry-Heilbronner Index, FBS = Fake Bad Scale. Raw scores are reported for each of the validity measures. Min. = minimum value. Max. = Maximum value.

16

A Study to Determine Homogeneity or Heterogeneity

Table 2. Correlations between MMPI-2 and WMT indices.

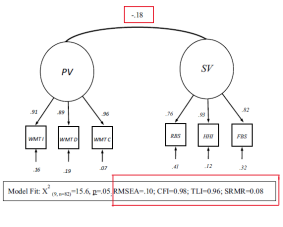
	RBS	HHI	FBS	WMT-I	WMT-D	WMT-C
RBS		.72**	.61**	−.31**	−.34**	−.36**
HHI			.77**	−.06	−.14	−.12
FBS				−.10	−.18	−.15
WMT-I					.82**	.88**
WMT-D						.86**
WMT-C						

RBS = Response Bias Scale, HHI = henry-Heilbronner Index, FBS = Fake Bad Scale, WMT = Word Memory Test, WMT-I = Word Memory Test Immediate Recognition, WMT-D = Word Memory Test-Delayed Recognition, WMT-C = Word Memory Test Recognition Consistency. N = 82. ** = p < .01.

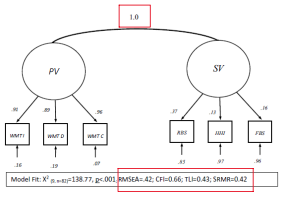
17

A Study to Determine Homogeneity or Heterogeneity

Free Estimation



Fixed Estimation of 1.0



18

A Study to Determine Homogeneity or Heterogeneity

- Conclusions
  - PVTs and SVTs do not represent a homogenous construct
  - Their relationship is trivial
- Implications
  - PVTs and SVTs are not interchangeable
    - Both constructs should be measured
  - These findings should be replicated with other instruments and in other populations

19

Recommendations

1. Control the risk of false positives
   - We should employ multiple validity indicators that possess high specificity and have defined base rate data
2. Mitigate the risk of test security violations
   - We should develop new PVTs that incorporate innovative methodology (e.g., biometrics, implicit memory)
3. Collect diverse normative samples
   - Norms for specific populations
   - Norms for diverse demographic groups
4. Measure PVTs *and* SVTs—they are not the same
5. Operationalize PVTs as measures of improbable performance and SVTs as indicators of improbable symptom reporting

20

QUESTIONS & ANSWERS

21

MAYO CLINIC

THANK YOU

22

Development of Symptom Validity Indices within Brief Psychological Symptom Inventories

Ryan W. Schroeder, PsyD, ABPP

AACN 2025

1

Employee Disclosure

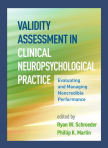
- I work for the Robert J Dole VA Medical Center (Wichita, KS)
  - The contents of research and any statements made during this presentation do not necessarily represent the views of the Robert J Dole VA Medical Center, the U.S. Department of Veterans Affairs, or the U.S. Government

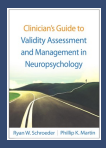
AACN 2025

2

Financial Disclosure

- I have a part-time forensic private practice
- I am a co-editor and co-author on validity-based books



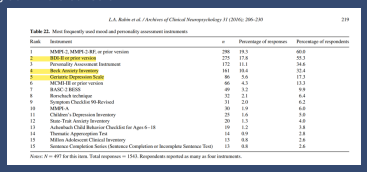


AACN 2025

3

Why Important to have Validity Scales in Brief Symptom Inventories?

1. Commonly administered tests



L.A. Rabin et al. / Archives of Clinical Neuropsychology 33 (2018): 206–230 219

Table 22. Most frequently used mood and personality assessment instruments

Rank	Instrument	*n*	Percentage of responses	Percentage of respondents
1	MMPI-2, MMPI-2-RF, or prior versions	298	19.5	60.0
2	BDI-II or prior versions	225	17.0	51.3
3	Personality Assessment Inventory	172	11.1	34.8
4	Beck Anxiety Inventory	161	10.4	32.4
5	Geriatric Depression Scale	86	5.6	17.3
6	MCMI-III or prior version	66	4.3	13.3
7	BASC-2 BESS	49	3.2	9.9
8	Rorschach technique	32	2.1	6.4
9	Symptom Checklist 90-Revised	31	2.0	6.2
10	MMPI-A	30	1.9	6.0
11	Children's Depression Inventory	23	1.6	5.0
12	Beck Youth Inventories	19	1.2	3.8
13	Achenbach Child Behavior Checklist for Ages 6–18	19	1.2	3.8
14	Thematic Apperception Test	14	0.9	2.8
15	Millon Adolescent Clinical Inventory	13	0.8	2.6
15	Sentence Completion Series (Sentence Completion or Incomplete Sentence Test)	13	0.8	2.6

Notes: $N = 497$ for this item. Total responses = 1543. Respondents reported as many as four instruments.

- If no validity scales included, the validity of responses is questionable
  - Results in the need to administer additional tests that include validity scales

AACN 2025

4

Why Important to have Validity Scales in Brief Symptom Inventories?

2. Other validity scales (F, Fp, NIM, MAL) often assess for overreported psychopathology more generally
   - MMPI-2-RF manual: "Elevated scores on F-r are associated with over-reporting of a broad range of psychological, cognitive and somatic symptoms" (p. 26)
   - PAI manual 2nd edition: "The Negative Impression (NIM) scale contains items that present an exaggerated unfavorable impression or represent extremely bizarre and unlikely symptoms" (p. 29)

- Arguably, validity scales from a symptom inventory might be better at determining whether a specific condition is being feigned

AACN 2025

5

Why Important to have Validity Scales in Brief Symptom Inventories?

3. Other providers might omit symptom validity indices
   - PCL-5 score of 71 → Claim "They have significant PTSD"
   - BDI-2 score 59 → Claim "They are very depressed"
     - Allows reviewers to have access to validity scales that would otherwise be absent

AACN 2025

6

Symptom Validity Scales within Symptom Inventories

PTSD Checklist for DSM-5 (PCL-5)

- Shura, R. D., Rowland, J. A., Miskey, H. M., Ord, A. S., Magnante, A. T., VA Mid-Atlantic MIRECC Workgroup, & Martindale, S. L. (2023). Symptom validity indices in the Posttraumatic Stress Disorder Checklist for DSM-5. *Journal of Traumatic Stress*, *36*(5), 919-931.
- Schroeder, R. W., & Bieu, R. K. (2024). Exploration of PCL-5 symptom validity indices for detection of exaggerated and feigned PTSD. *Journal of Clinical and Experimental Neuropsychology*, *46*(2), 152-161.
- Schroeder, R. W., Spector, J., Snodgrass, M., & Bieu, R. K. (in press). Validation of PCL-5 symptom validity indices in a cross-cultural forensic sample. *Journal of Clinical and Experimental Neuropsychology*.

Beck Depression Inventory-2 (BDI-2)

- Fuermaier, A. B., Dandachi-Fitzgerald, B., & Lehrer, J. (2023). Validity assessment of early retirement claimants: Symptom overreporting on the Beck Depression Inventory–II. *Applied Neuropsychology: Adult*, 1-7.
- Shura, R. D., Schroeder, R. W., Ord, A. S., Bieu, R. K., O'Connor, V. L., Magnante, A. T., ... & Rowland, J. A. (2024). Symptom validity indices for the Beck Depression Inventory-II: development and cross-validation in research and clinical samples. *The Clinical Neuropsychologist*, 1-19.
- Boucher, C. M., Giromini, L., Roth, R. M., & Erdodi, L. A. (2024). The Beck Depression Inventory—Second Edition as a Symptom Validity Test: Importing European Cutoffs to the USA. *Psychological Injury and Law*, 1-12.
- Merten, T. (2024). Highly Elevated Scores on the Beck Depression Inventory–Second Edition as an Indicator of Noncredible Symptom Report. *Assessment*. 10731911241304214.

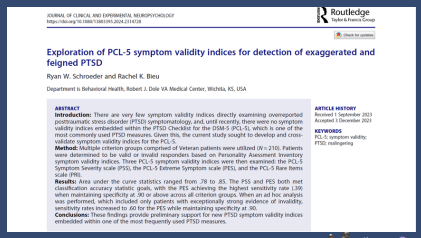
Beck Anxiety Inventory (BAI)

- Snodgrass, M. A., Bieu, R. K., & Schroeder, R. W. (2024). Development of a symptom validity index for the beck anxiety inventory. *The Clinical Neuropsychologist*, 1-16.

AACN 2025

7

Symptom Validity Indices within the PCL-5



JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY
https://doi.org/10.1080/13803395.2024.2314728

Routledge
Taylor & Francis Group

Check for updates

Exploration of PCL-5 symptom validity indices for detection of exaggerated and feigned PTSD

Ryan W. Schroeder and Rachel K. Bieu

Department is Behavioral Health, Robert J. Dole VA Medical Center, Wichita, KS, USA

ABSTRACT

Introduction: There are very few symptom validity indices directly examining overreported posttraumatic stress disorder (PTSD) symptomatology, and, until recently, there were no symptom validity indices embedded within the PTSD Checklist for the DSM-5 (PCL-5), which is one of the most commonly used PTSD measures. Given this, the current study sought to develop and cross-validate symptom validity indices for the PCL-5.

Method: Multiple criterion groups comprised of Veteran patients were utilized (*N* = 210). Patients were determined to be valid or invalid responders based on Personality Assessment Inventory symptom validity indices. Three PCL-5 symptom validity indices were then examined: the PCL-5 Symptom Severity scale (PSS), the PCL-5 Extreme Symptom scale (PES), and the PCL-5 Rare Items scale (PRS).

Results: Area under the curve statistics ranged from .78 to .85. The PSS and PES both met classification accuracy statistic goals, with the PES achieving the highest sensitivity rate (.39) when maintaining specificity at .90 or above across all criterion groups. When an ad hoc analysis was performed, which included only patients with exceptionally strong evidence of invalidity, sensitivity rates increased to .60 for the PES while maintaining specificity at .90.

Conclusions: These findings provide preliminary support for new PTSD symptom validity indices embedded within one of the most frequently used PTSD measures.

ARTICLE HISTORY
Received 1 September 2023
Accepted 3 December 2023

KEYWORDS
PCL-5; symptom validity; PTSD; malingering

AACN 2025

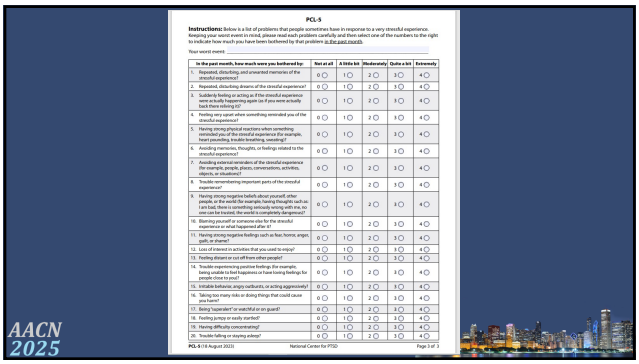
8

PTSD Checklist for DSM-5 (PCL-5)

- 20 Item checklist
  - The 20 items are the 20 DSM-5 symptoms
- Rate each item/symptom based on how bothered examinees have been by the symptom over the last month
  - Symptoms are rated on a scale of 0-4
    - Not at all (0), A little bit (1), Moderately (2), Quite a bit (3), Extremely (4)
- Score range: 0-80
- Scores in 31-33 range suggest clinically significant PTSD

AACN 2025

9



10

3 Symptom Validity Indices within the PCL-5

1. PCL-5 Symptom Severity scale (PSS)
   - Concurrently developed by Shura et al. and Schroeder & Bieu
   - Scoring → PCL-5 total score (0-80)
   - "This index was created based on the premise that individuals who exaggerate or feign PTSD will overreport the total severity of PTSD symptomatology as compared to individuals with genuine PTSD." (Schroeder & Bieu, 2024)

AACN 2025

11

3 Symptom Validity Indices within the PCL-5

2. PCL-5 Extreme Symptom scale (PES)
   - Developed by Schroeder & Bieu
   - Scoring → Summing the number of items rated as 3 or 4 (0-20)
   - "This index was based on the premise that individuals who exaggerate or feign PTSD might endorse a greater number of immoderate scores as compared to individuals with genuine PTSD." (Schroeder & Bieu, 2024)

AACN 2025

12

3 Symptom Validity Indices within the PCL-5

3. PCL-5 Rare Items scale (PRI)
   - Developed by Shura et al.
   - Derived by identifying the PCL-5 items endorsed:
     - (1) by less than 10% of a diagnostically diverse post-deployment sample AND
     - (2) by less than 20% of the PTSD subsample
   - Six items identified:
     - Items 8 and 16 endorsed as either "3" or "4"
     - Items 3, 9, 10, and 15 endorsed "4"
   - Created based on "items that were rarely endorsed at high levels by veterans and quasirarely endorsed by individuals with PTSD who responded validly." (Shura et al., 2023)

AACN 2025

13

How Did We Examine These 3 Scales?

- Sample of 210 clinically evaluated veterans
- Grouped:
  - Validly reporting symptoms based on PAI validity scales
  - Invalidly reporting symptoms based on PAI validity scales
- Valid group further analyzed:
  - Total sample (*n*=187)
  - No history of PTSD – but often history of other psychiatric disorders (*n*=60)
  - A history of PTSD in medical records but not currently meeting full DSM-5 criteria (*n*=30)
  - Currently meeting DSM-5 criteria for PTSD (*n*=97)

AACN 2025

14

Group Data

Table 1. Basic demographic and background information for patients.

Group	n	Age	Education	Gender (%Male)	Race (%White)	PCL-5 Score
Full Sample	210	42.86 (12.57)	14.10 (2.21)	77%	87%	36.17 (19.29)
Invalid	23	42.52 (13.12)	14.61 (2.25)	61%	64%	55.83 (11.75)
Valid Full Sample	187	42.89 (12.54)	14.04 (2.21)	77%	88%	33.75 (18.66)
Valid Non-PTSD	60	43.68 (13.14)	13.95 (2.14)	82%	90%	15.87 (10.67)
Valid Hx-PTSD	30	41.97 (11.67)	14.10 (2.06)	77%	100%	23.60 (8.77)
Valid Active PTSD	97	42.70 (12.52)	14.07 (2.31)	74%	83%	47.95 (12.07)

Note: PCL-5 = PTSD Checklist for DSM-5; PTSD = Posttraumatic Stress Disorder; Valid Hx-PTSD = Valid Historical-PTSD.

AACN 2025

15

ROC Analyses

- Ran ROC analyses

Table 2. Area under the curve, sensitivity, and specificity for the PCL-5 symptom validity scales.

Symptom Validity Scale	AUC	Sensitivity	Specificity
PCL-5 Symptom Severity scale	0.83	0.48	0.90
PCL-5 Extreme Symptom scale	0.85	0.44	0.92
PCL-5 Rare Items scale	0.78	0.26	0.90

Note: PCL-5 = PTSD Checklist for DSM-5; AUC = Area under the curve; Sensitivity and specificity rates are based on the first cutoff that is associated with a specificity rate of 90% or better in the full sample.

AACN 2025

16

PSS Findings

Table 3. Sensitivity and specificity rates for the PCL-5 symptom severity scale.

Group:	Invalid	Valid Full Sample	Valid Non-PTSD	Valid Historical PTSD	Valid Active PTSD
PCL-5 score ≥	Sensitivity	Specificity	Specificity	Specificity	Specificity
50	0.70	0.81	1.00	1.00	0.64
51	0.65	0.82	1.00	1.00	0.65
52	0.61	0.82	1.00	1.00	0.66
53	0.57	0.83	1.00	1.00	0.68
54	0.57	0.84	1.00	1.00	0.69
55	0.57	0.85	1.00	1.00	0.71
56	0.57	0.86	1.00	1.00	0.72
57	0.57	0.87	1.00	1.00	0.74
58	0.48	0.88	1.00	1.00	0.76
59	0.48	0.89	1.00	1.00	0.79
60	0.48	0.90	1.00	1.00	0.81
61	0.48	0.93	1.00	1.00	0.86
62	0.44	0.93	1.00	1.00	0.87
63	0.39	0.93	1.00	1.00	0.87
64	**0.30**	**0.95**	**1.00**	**1.00**	**0.90**
65	0.30	0.95	1.00	1.00	0.90
66	0.26	0.95	1.00	1.00	0.91
67	0.22	0.96	1.00	1.00	0.93
68	0.13	0.96	1.00	1.00	0.93
69	0.13	0.97	1.00	1.00	0.94
70	0.09	0.97	1.00	1.00	0.95
71	0.09	0.97	1.00	1.00	0.95
72	0.04	0.97	1.00	1.00	0.95
73	0.04	0.97	1.00	1.00	0.95
74	0.04	0.99	1.00	1.00	0.98
75	0.00	0.99	1.00	1.00	0.99
76	0.00	1.00	1.00	1.00	1.00

Note: PCL-5 = PTSD Checklist for DSM-5.

AACN 2025

17

Clinical Question

- "Does my examinee's score represent PTSD or exaggeration/feigning of PTSD?"

AACN 2025

18

PSS Findings

Table 3. Sensitivity and specificity rates for the PCL-5 symptom severity scale.

Group:	Invalid	Valid Full Sample	Valid Non-PTSD	Valid Historical-PTSD	Valid Active PTSD
PCL-5 score ≥	Sensitivity	Specificity	Specificity	Specificity	Specificity
50	0.70	0.81	1.00	1.00	0.64
51	0.65	0.82	1.00	1.00	0.65
52	0.61	0.82	1.00	1.00	0.66
53	0.57	0.83	1.00	1.00	0.68
54	0.57	0.84	1.00	1.00	0.69
55	0.57	0.85	1.00	1.00	0.71
56	0.57	0.86	1.00	1.00	0.72
57	0.57	0.87	1.00	1.00	0.74
58	0.48	0.88	1.00	1.00	0.76
59	0.48	0.89	1.00	1.00	0.79
60	0.48	0.90	1.00	1.00	0.81
61	0.48	0.93	1.00	1.00	0.86
62	0.44	0.93	1.00	1.00	0.87
63	0.39	0.93	1.00	1.00	0.87
64	**0.30**	**0.95**	**1.00**	**1.00**	**0.90**
65	0.30	0.95	1.00	1.00	0.90
66	0.26	0.95	1.00	1.00	0.91
67	0.22	0.96	1.00	1.00	0.93
68	0.13	0.96	1.00	1.00	0.93
69	0.13	0.97	1.00	1.00	0.94
70	0.09	0.97	1.00	1.00	0.95
71	0.09	0.97	1.00	1.00	0.95
72	0.04	0.97	1.00	1.00	0.95
73	0.04	0.97	1.00	1.00	0.95
74	0.04	0.99	1.00	1.00	0.98
75	0.00	0.99	1.00	1.00	0.99
76	0.00	1.00	1.00	1.00	1.00

Note: PCL-5 = PTSD Checklist for DSM-5.

AACN 2025

19

PSS Findings

Symptom Inventory
Contain true clinical items

vs

MMPI/PAI/Etc.
Contain atypical items

AACN 2025

20

PSS Findings

Table 3. Sensitivity and specificity rates for the PCL-5 symptom severity scale.

Group:	Invalid	Valid Full Sample	Valid Non-PTSD	Valid Historical-PTSD	Valid Active PTSD
PCL-5 score ≥	Sensitivity	Specificity	Specificity	Specificity	Specificity
50	0.70	0.81	1.00	1.00	0.64
51	0.65	0.82	1.00	1.00	0.65
52	0.61	0.82	1.00	1.00	0.66
53	0.57	0.83	1.00	1.00	0.68
54	0.57	0.84	1.00	1.00	0.69
55	0.57	0.85	1.00	1.00	0.71
56	0.57	0.86	1.00	1.00	0.72
57	0.57	0.87	1.00	1.00	0.74
58	0.48	0.88	1.00	1.00	0.76
59	0.48	0.89	1.00	1.00	0.79
60	0.48	0.90	1.00	1.00	0.81
61	0.48	0.93	1.00	1.00	0.86
62	0.44	0.93	1.00	1.00	0.87
63	0.39	0.93	1.00	1.00	0.87
64	**0.30**	**0.95**	**1.00**	**1.00**	**0.90**
65	0.30	0.95	1.00	1.00	0.90
66	0.26	0.95	1.00	1.00	0.91
67	0.22	0.96	1.00	1.00	0.93
68	0.13	0.96	1.00	1.00	0.93
69	0.13	0.97	1.00	1.00	0.94
70	0.09	0.97	1.00	1.00	0.95
71	0.09	0.97	1.00	1.00	0.95
72	0.04	0.97	1.00	1.00	0.95
73	0.04	0.97	1.00	1.00	0.95
74	0.04	0.99	1.00	1.00	0.98
75	0.00	0.99	1.00	1.00	0.99
76	0.00	1.00	1.00	1.00	1.00

Note: PCL-5 = PTSD Checklist for DSM-5.

AACN 2025

21

PES Findings

Table 5. Sensitivity and specificity rates for the PCL-5 extreme symptom scale

Group:	Invalid	Valid Full Sample	Valid Non-PTSD	Valid Historical-PTSD	Valid Active PTSD
Items ≥	Sensitivity	Specificity	Specificity	Specificity	Specificity
1	1.00	0.18	0.47	0.17	0.00
2	1.00	0.27	0.62	0.47	0.00
3	1.00	0.34	0.75	0.53	0.02
4	1.00	0.40	0.22	0.63	0.06
5	1.00	0.49	0.88	0.80	0.15
6	0.96	0.53	0.92	0.83	0.20
7	0.91	0.58	0.93	0.90	0.26
8	0.91	0.61	0.98	0.93	0.29
9	0.87	0.65	0.98	0.97	0.34
10	0.83	0.70	1.00	1.00	0.42
11	0.83	0.76	1.00	1.00	0.55
12	0.74	0.81	1.00	1.00	0.64
13	0.65	0.82	1.00	1.00	0.66
14	0.65	0.84	1.00	1.00	0.69
15	0.52	0.88	1.00	1.00	0.77
16	0.44	0.92	1.00	1.00	0.84
17	**0.39**	**0.95**	**1.00**	**1.00**	**0.90**
18	0.17	0.96	1.00	1.00	0.93
19	0.04	0.98	1.00	1.00	0.96
20	0.00	1.00	1.00	1.00	1.00

Note: PCL-5 = PTSD Checklist for DSM-5.

AACN 2025

22

PRI Findings

Table 4. Sensitivity and specificity rates for the PCL-5 rare item scale.

Group:	Invalid	Valid Full Sample	Valid Non-PTSD	Valid Historical-PTSD	Valid Active PTSD
Items ≥	Sensitivity	Specificity	Specificity	Specificity	Specificity
1	0.87	0.65	0.93	0.97	0.37
2	0.52	0.86	1.00	1.00	0.72
3	0.26	0.91	1.00	1.00	0.84
4	**0.13**	**0.96**	**1.00**	**1.00**	**0.92**
5	0.00	0.99	1.00	1.00	0.98
6	0.00	1.00	1.00	1.00	1.00

Note: PCL-5 = PTSD Checklist for DSM-5.

AACN 2025

23

New Cross-Validation Data

- Schroeder et al. (in press) → Validation of PCL-5 symptom validity indices in a cross-cultural forensic sample
  - PCL-5, IOP-29, and other measures administered to Balkan individuals
    - Contractors previously employed at US military bases in Afghanistan and Iraq
    - All claimed PTSD and were undergoing federal WC evaluations related to the claims
  - Using IOP-29 as the criterion SVT:
    - First cutoff for PSS to result in 90+% specificity was 63
      - Sens = 47%, Spec = 90%
    - First cutoff for PES to result in 90+% specificity was 17
      - Sens = 33%, Spec = 92%
    - First cutoff for PRI to result in 90+% specificity was 3
      - Sens = 43%, Spec = 94%

AACN 2025

24

Additional Published Data

- Shura et al. (2023)
  - PSS cutoff of ≥64 → Exact classification accuracy rates not reported at cutoff but specificity >90%
  - PRI cutoff of ≥4 → Sens: 10-19%, Spec: 98-100%
  - PES not examined

AACN 2025

25

Cutoffs Maintaining 90+% Specificity Across All Studies

- PSS = 64+
  - Shura et al. (2023)
  - Schroeder & Bieu (2024)
  - Schroeder et al. (in press)
- PES = 17+
  - Schroeder & Bieu (2024)
  - Schroeder et al. (in press)
- PRI = 4+
  - Shura et al. (2023)
  - Schroeder & Bieu (2024)
  - Schroeder et al. (in press)

AACN 2025

26

PSS Critique

- "The PSS is derived from the PCL-5 total score, which is based on true PTSD symptomatology. Why are you trying to say that the examinee exaggerated his symptoms when he was just reporting true DSM-5 symptoms?"

AACN 2025

27

PSS Critique Response

- The cutoff for invalidity is unrealistically high for outpatients
  - Essentially double the clinical cutoff (31-33) for PTSD
  AND
  - It equates to endorsing 80% of the maximum obtainable score on the PCL-5 (i.e., 64/80 = .80)
- This would represent a profound degree of pathology, that is uncommon in outpatients with true PTSD

AACN 2025

28

PSS Critique Response

Table 7. Average PCL-5 scores across studies of individuals with PTSD.

Study	Participants	Sample Size	Mean PCL-5 Score	PCL-5 Score 1.3 SD Above Mean
Wortmann et al. (2016)	Service Members; 70% met PTSD criteria	912	42	62
Marx et al. (2022)	Veterans; PTSD present in all	198	48	65
Marx et al. (2022)	Veterans; PTSD present in all	119	48	69
Miskey et al. (2020)	Veterans; PTSD present in all	110	47	67
Peterson et al. (2022)	Active Duty & Veterans; PTSD present in all	120	50	68
Davis et al. (2020)	Individuals with PTSD; 90% Veterans	108	45	61
Davis et al. (2020)	Individuals with PTSD; 93% Veterans	101	45	62
All studies combined	All of the above	1668	46.43	64.86

Note: PCL-5 scores are rounded to the nearest whole number; PCL-5 = PTSD Checklist for DSM-5; PTSD = Posttraumatic Stress Disorder; SD = Standard deviation.

AACN 2025

29

PSS Critique Response

- Specificity rates 90+% at 64+ cutoff per:
  - Shura et al. (2023)
  - Schroeder & Bieu (2024)
  - Schroeder et al. (in press)

AACN 2025

30

PES Critique

- "Wouldn't you expect someone with PTSD to endorse having high scores on a PTSD symptom checklist? Why are you claiming her responses are invalid?"

AACN 2025

31

PES Critique Response

- The cutoff for invalidity is extremely high
  - A score of 17 indicates that 85% (i.e., 17/20 = .85) of all PCL-5 items were endorsed at the most severe ratings (i.e., 3 or 4).
- Research indicates that it is uncommon for outpatients with PTSD to be highly and essentially equally distressed by nearly every possible symptom listed in the DSM-5
  - Bovin et al. (2016) found:
    - Veterans had average PCL-5 scores of "1" (i.e., "a little bit" distress) on 12/20 items
    - They had average scores of "2" (i.e., moderately distressing) on the remaining items
- Likewise, data from Schroeder & Bieu (2024) & Schroeder et al. (in press) show that scores in the 17+ range were more likely than not to be associated with invalidity (also a 90% specificity rate).

AACN 2025

32

PRI Critique

- "You're saying that my examinee is feigning because he endorsed a couple high scores on this checklist derived of true PTSD symptomatology?"

AACN 2025

33

PRI Critique Response

- "While the items do represent true PTSD symptomatology, these specific items are not commonly endorsed at high intensity ratings. This is supported by Shura et al.'s PTSD sample, Schroeder and Bieu's PTSD sample, and Schroeder et al.'s PTSD sample. Even research outside of SVT research shows that the items comprising the PRI are some of the items with the lost mean responses on the PCL-5 (Bovin et al., 2016). Overall, the research has consistently shown that a cutoff of 4 or more on this scale is indicative of exaggerated or feigned pathology in outpatient PTSD samples."

AACN 2025

34

Final Points

- PCL-5 is a free inventory, and there are now well validated symptom validity indices for it
- Validity scales add to the minimal number of PTSD-specific SVTs
- Caution should be utilized in some situations, though
  - Ongoing research with different cultural groups is warranted
  - Ongoing research with individuals who have other types of trauma, particularly sexual and physical violence, would be beneficial
  - Unclear whether the cutoffs noted in the study would be appropriate for inpatients with PTSD

AACN 2025

35

Final Points

- PSS = 64+
  - Shura et al. (2023)
  - Schroeder & Bieu (2024)
  - Schroeder et al. (in press)
- PES = 17+
  - Schroeder & Bieu (2024)
  - Schroeder et al. (in press)
- PRI = 4+
  - Shura et al. (2023)
  - Schroeder & Bieu (2024)
  - Schroeder et al. (in press)

AACN 2025

36

References

- Boucher, C. M., Giromini, L., Roth, R. M., & Erdodi, L. A. (2024). The Beck Depression Inventory—Second Edition as a Symptom Validity Test: Importing European Cutoffs to the USA. *Psychological Injury and Law*, 1-12.
- Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders–fifth edition (PCL-5) in veterans. *Psychological assessment*, *28*(11), 1379.
- Fuermaier, A. B., Dandachi-Fitzgerald, B., & Lehrner, J. (2023). Validity assessment of early retirement claimants: Symptom overreporting on the Beck Depression Inventory–II. *Applied Neuropsychology: Adult*, 1-7.
- Merten, T. (2024). Highly Elevated Scores on the Beck Depression Inventory–Second Edition as an Indicator of Noncredible Symptom Report. *Assessment*, 10731911241304214.
- Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology*, *31*(3), 206-230.
- Schroeder, R. W., & Bieu, R. K. (2024). Exploration of PCL-5 symptom validity indices for detection of exaggerated and feigned PTSD. *Journal of Clinical and Experimental Neuropsychology*, *46*(2), 152-161.
- Schroeder, R. W., Spector, J., Snodgrass, M., & Bieu, R. K. (in press). Validation of PCL-5 symptom validity indices in a cross-cultural forensic sample. *Journal of Clinical and Experimental Neuropsychology*.
- Shura, R. D., Rowland, J. A., Miskey, H. M., Ord, A. S., Magnante, A. T., VA Mid-Atlantic MIRECC Workgroup, & Martindale, S. L. (2023). Symptom validity indices in the Posttraumatic Stress Disorder Checklist for DSM-5. *Journal of Traumatic Stress*, *36*(5), 919-931.
- Shura, R. D., Schroeder, R. W., Ord, A. S., Bieu, R. K., O'Connor, V. L., Magnante, A. T., ... & Rowland, J. A. (2024). Symptom validity indices for the Beck Depression Inventory-II: development and cross-validation in research and clinical samples. *The Clinical Neuropsychologist*, 1-19.
- Snodgrass, M. A., Bieu, R. K., & Schroeder, R. W. (2024). Development of a symptom validity index for the beck anxiety inventory. *The Clinical Neuropsychologist*, 1-16.

AACN 2025

37

Thank You!

- Ryan Schroeder
- Ryan.w.Schroeder.psyd@hotmail.com

AACN 2025

38

Embedded PVTs: Future Directions for Research and Evidence-Based Clinical Practice

Jason R. Soble, Ph.D., ABPP
Associate Professor of Clinical Psychiatry & Neurology
University of Illinois Chicago College of Medicine

AACN 2025

1

Disclosures

- Any opinions, findings, conclusions or recommendations expressed are those of the presenter.
- No conflicts of interest to disclose

AACN 2025

2

Current State of PVTs in Neuropsychology

- PVTs development has bourgeoned over the past 2 decades (Schroeder & Martin, 2021; Soble et al., 2021; Sweet et al., 2021)
- Why? – Need and Demand
  - Neuropsychologists reporting routine PVT usage:
    - 2007: 56% (Sharland & Gfeller, 2007)
    - 2015: 92% (Martin et al., 2015)
    - 2024: 93%-100% (Martin et al., 2024)
  - Research and development of freestanding and embedded measures
    - Between 1990-2007: >300 publications on the topic of PVTs (Boone, 2007)
    - Between 2007-2015: >1,400 publications (Boone, 2021)
  - Potential for invalidity is omnipresent in neuropsychological evaluations
    - Forensic exams: 40-50%; Clinical exams 15% median (range: 5-50%) (Martin & Schroeder, 2020)

AACN 2025

3

The PVT Balancing Act

1. Time/Cost Burden
2. Test Security
3. Accuracy
   - Individual PVTs
   - Invalidity determination across multiple PVTs

Soble et al., (2025)

AACN 2025

4

Innovations in Assessing Performance and Symptom Validity- JCEN (2024)

- Peak, A. M., Marceaux, J. C., Chicota-Carroll, C., & Soble, J. R. (2024). Cross-validation of the Trail Making Test as a non-memory-based embedded performance validity test among veterans with and without cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, *46*(1), 16–24.
- Attempted to replicate earlier TMT PVT findings reported in a mixed neuropsychiatric civilian population (White et al., 2020)

AACN 2025

5

Peak et al. (2024)

Primary Diagnosis	Valid-Unimpaired *n* = 40	Valid-Impaired *n* = 35	Invalid *n* = 25
Alcohol Use Disorder	-	1	-
Alzheimer's Disease	-	2	1
Amnestic Mild Cognitive Impairment	-	4	-
Anxiety	4	-	1
Attention-Deficit/Hyperactivity Disorder	5	-	-
Autoimmune Condition	1	-	-
Bipolar Disorder	1	-	-
Cerebrovascular Disease/Stroke	-	14	1
Chronic Pain	2	-	-
Depression	5	-	3
Epilepsy/Seizures	1	5	1
Frontotemporal Degeneration	-	2	-
Malingering (rule-out/suspected)	-	-	5
Multiple Comorbid Etiologies	-	2	1
No Diagnosis	7	-	-
Parkinson's Disease	-	1	-
Posttraumatic Stress Disorder	5	-	7
Psychotic Disorder	1	-	-
Sleep Disorder	2	-	-
Somatic Symptom Disorder	1	-	-
Substance Use Disorder	1	1	-
Traumatic Brain Injury (Mild/Concussion)	2	-	3
Traumatic Brain Injury (Mild-Complicated)	1	-	1
Traumatic Brain Injury (Moderate-Severe)	1	2	-
Tumor	-	1	1

AACN 2025

6

Peak et al. (2024)-Overall Sample

TMT score	AUC	Cutoff	SN	SP	White et al. (2020)	
TMT-A raw	.72***	≥55	.44	.85	.72**	≥56; .35/.90
		≥62	**.40**	**.93**		
		≥64	.36	.95		
TMT-A T-score	.75***	≤32	.44	.92	.72**	≤34; .35/.89
		≤33	**.48**	**.92**		
		≤36	.52	.87		
TMT-B raw	.74***	≥199	.36	.88	.74***	≥285; .27/.89
		≥220	**.32**	**.91**		
		≥234	.28	.91		
TMT-B T-score	.78***	≤29	.36	.93	.73***	≤27; .42/.90
		≤30	**.40**	**.92**		
		≤32	.44	.88		

AACN 2025

7

Peak et al. (2024)-Unimpaired Subsample

TMT score	AUC	Cutoff	SN	SP	White et al. (2020)	
TMT-A raw	.80***	≥47	.60	.87	.88***	≥39; .58/.94
		≥49	**.56**	**.90**		
		≥53	.48	.92		
TMT-A T-score	.81***	≤37	.52	.95	.80***	≤40; .58/.90
		≤40	**.60**	**.90**		
		≤41	.64	.87		
TMT-B raw	.86***	≥100	.72	.85	.94***	≥80; .85/.90
		≥109	**.72**	**.90**		
		≥119	.68	.92		
TMT-B T-score	.88***	≤38	.64	.95	.87***	≤40; .62/.90
		≤40	**.72**	**.90**		
		≤42	.80	.85		

AACN 2025

8

Peak et al. (2024)-Impaired Subsample

TMT score	AUC	Cutoff	SN	SP	White et al. (2020)	
TMT-A raw	.64	-	-	-	.65*	≥62; .31/.90
TMT-A T-score	.68*	≤28	.36	.91	.68**	≤31; .27/.90
		≤30	**.44**	**.91**		
		≤33	.48	.86		
TMT-B raw	.60	-	-	-	.65*	≥285; .27/.84
TMT-B T-score	.66*	≤25	.20	.94	.67*	≤21; .19/.90
		≤26	**.28**	**.94**		
		≤28	.32	.86		

AACN 2025

9

Validating Embedded PVTs- Research Study Design

- Simulation Studies
- Differential Prevalence Studies
- Criterion Grouping Studies (known-groups designs)

AACN 2025

10

Research Study Design

- **Simulation Studies**
  - Useful when no objective external grouping criterion exists
  - Relies on artificially constructed validity groups
  - Inflates classification accuracy
  - More limited generalizability to actual clinical/forensic examinees

AACN 2025

11

Research Study Design

- **Differential Prevalence Studies:**
  - Validity groups constructed using non-performance-based criteria (e.g., compensation seeking vs. no compensation seeking)
  - Assumes all in the invalid groups are invalid and vice versa without objective verification
  - Estimates of invalidity among compensation-seeking examinees typically plateau at ~40-50% (Chafetz & Underhill, 2013; Larrabee et al., 2009; Young et al., 2016)
  - Invalidity is not uncommon in non-forensic clinical contexts (median BR: 15%; range 5%-50%) (Martin & Schroeder, 2020)

AACN 2025

12

Research Study Design

- **Criterion Grouping Design:**
  - Currently recommend best practice for validating/cross-validating PVTs (Sweet et al., 2021)
  - Validity groups are constructed using objective, performance-based criteria
    - Minimizes need for subjective judgement and examinee status-based criteria
  - Previously validated PVTs that are independent of the test PVT(s) being examined are used establish validity groups
  - Test PVT(s) are then analyzed to determine classification accuracy for detecting invalidity and relevant psychometric properties

AACN 2025

13

Not all Criterion Grouping Studies are Created Equal

- Considerable heterogeneity exists in terms of both the quantity and quality of independent PVTs used across studies
- This considerable heterogeneity partially (though not exclusively) drives some of the variable (and in some cases) drastically discrepant findings across individual studies looking at the same embedded PVTs.
- Is *quantity* or *quality* of PVTs used the more critical consideration?

AACN 2025

14

Implications of the Number of PVTs used to Establish Validity Groups in Criterion Grouping Studies

- 1 PVT approach
  - Classification accuracy influenced heavily by strength of relationship between criterion and test PVT (Schroeder et al., 2019)
  - High risk for both false negative and false positive classification
- 6 PVT protocol: WMT, TOMM, WCT, DCT, RDS, RFIT (Soble et al., 2020)
  - All 2- and 3-PVT combinations compared to final validity status across the 6
  - 2-PVT combinations
    - Invalid: 17%-76% failed 2/2 and would be correctly classified as invalid
    - Invalid: 0%-38% failed 0/2 and would be incorrectly classified as valid (false negative)
    - Valid: 72%-90% failed 0/2 and would be correctly classified as valid
    - Valid: 0%-4% failed 2/2 and would be incorrectly classified as invalid (false positive)

AACN 2025

15

Implications of the Number of PVTs used to Establish Validity Groups in Criterion Grouping Studies

- 6 PVT protocol: WMT, TOMM, WCT, DCT, RDS, RFIT (Soble et al., 2020)
  - 3-PVT combinations
    - Invalid: 55%-97% failed ≥2/3 and would be correctly classified as invalid
    - Invalid: 0%-14% failed 0/3 and would be incorrectly classified as valid (false negative)
      - 0% for 9/18 and ≤6% for 6/18 3-PVT combinations
    - Valid: 66%-85% failed 0/3 and would be correctly classified as valid
    - Valid: 0%-6% failed ≥2/3 and would be incorrectly classified as invalid (false positive)
- Increasing the number of independent criterion PVTs results in more accurately defined validity groups, particularly for the invalid group

AACN 2025

16

Implications of Quality of PVTs used to Establish Validity Groups in Criterion Grouping Studies

- Increasing PVT quantity from 2-3 substantially dropped false negative rate, but quality of PVTs in the combination still mattered
  - Sensitivity of individual PVTs varies considerably
  - Adding more low quality (i.e., weak sensitivity) PVTs will not improve study internal validity and is less likely to yield replicable/generalizable results
- In short, both quantity and quality of independent PVTs are key!
- Correlations between criterion and test PVT(s) is also critical
  - Highly correlated measures often provide redundant information
  - No universally accepted standard, but ≤.50 is a common rule of thumb

AACN 2025

17

A Final Consideration for Criterion-Grouping Designs: Setting the Threshold for Invalidity

- ≥2 independent PVT failures is generally accepted invalidity threshold
  - If ≥90% specificity, this would be a rare event to occur by chance
  - (.01 or 1/10 x 1/10)
- Is this ≥2 PVT failure invalidity threshold invariant irrespective of the number of PVTs administered?
  - For freestanding PVTs, probably yes
  - **For embedded PVTs, no**

AACN 2025

18

Setting the Threshold for Invalidity for Embedded PVT Validation Studies

- ≥2 independent PVT failures is the optimal threshold if 4-9 total PVTs are administered (e.g., Davis & Millis, 2014; Delorla et al., 2023; Larrabee, 2014; Martin et al., 2022; Pearson, 2009; Victor et al., 2009)
- However, when 10 embedded PVTs are administered:
  - ≥2 failures (SN 77%/SP 68%); ≥3 failures (SN 61%/SP 90%) (Larrabee et al., 2019)
  - ≥2 failures (SN 86%/SP 76%); ≥3 failures (SN 69%/SP 92%) (Leese et al., 2025)
- Adjustment of threshold used to determine invalidity from ≥2 fails to ≥3 failures is needed if ≥10 criterion embedded PVTs are used

AACN 2025

19

Validating Embedded PVTs: Best Practices and Future Directions Summary

1. Studies using a criterion-grouping methodology are preferred and should be given the most weight
2. No universal optimal number of independent criterion PVTs exists
   - Existing research shows serious limitations with using only 1-2 PVTs
   - Clinical practice median: 5 PVTs (2 freestanding; 3 embedded) (Martin et al., 2015)
3. Quality of criterion PVTs also matters a lot
   - Using poorly-validated or insensitive measures as criterion PVTs does not magically yield better results
4. Correlations between test and criterion PVTs should be reported and carefully examined

AACN 2025

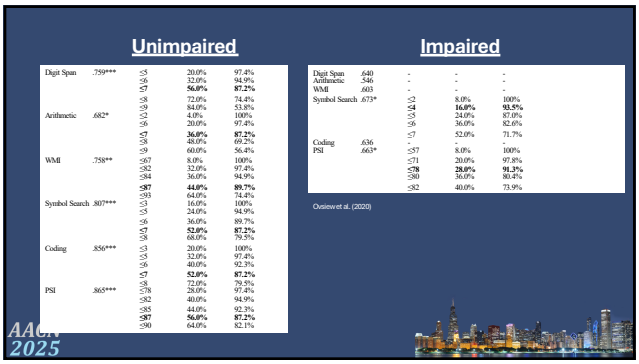
20

Validating Embedded PVTs- Issues Related to Clinical Population and PVT Paradigm

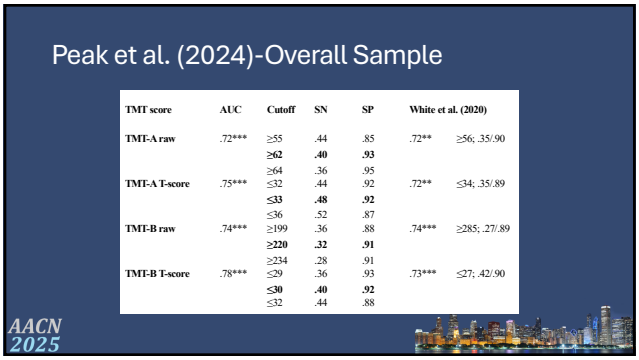
- Research samples/clinical populations in which PVTs are validated can significantly influence resulting diagnostic and psychometric properties (for better or worse)
- Principle is relevant for all PVTs, but especially embedded measures
- Embedded PVTs validated exclusively in populations with no to minimal impairment are likely to have far less utility

AACN 2025

21

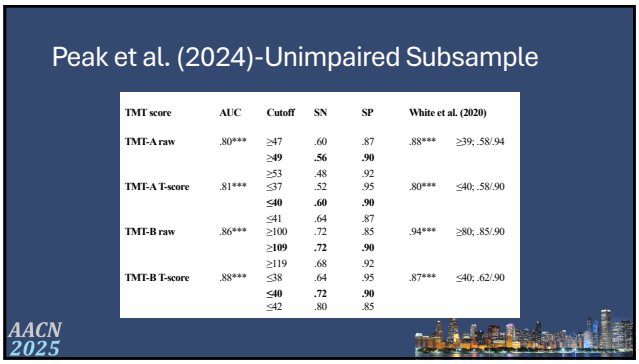


22



TMT score	AUC	Cutoff	SN	SP	White et al. (2020)	
TMT-A raw	.72***	≥55	.44	.85	.72**	≥56; .35/.90
		≥62	**.40**	**.93**		
		≥64	.36	.95		
TMT-A T-score	.75***	≤32	.44	.92	.72**	≤34; .35/.89
		≤33	**.48**	**.92**		
		≤36	.52	.87		
TMT-B raw	.74***	≥199	.36	.88	.74***	≥285; .27/.89
		≥220	**.32**	**.91**		
		≥234	.28	.91		
TMT-B T-score	.78***	≤29	.36	.93	.73***	≤27; .42/.90
		≤30	**.40**	**.92**		
		≤32	.44	.88		

23



TMT score	AUC	Cutoff	SN	SP	White et al. (2020)	
TMT-A raw	.80***	≥47	.60	.87	.88***	≥39; .58/.94
		≥49	**.56**	**.90**		
		≥53	.48	.92		
TMT-A T-score	.81***	≤37	.52	.95	.80***	≤40; .58/.90
		≤40	**.60**	**.90**		
		≤41	.64	.87		
TMT-B raw	.86***	≥100	.72	.85	.94***	≥80; .85/.90
		≥109	**.72**	**.90**		
		≥119	.68	.92		
TMT-B T-score	.88***	≤38	.64	.95	.87***	≤40; .62/.90
		≤40	**.72**	**.90**		
		≤42	.80	.85		

24

Peak et al. (2024)-Impaired Subsample

TMT score	AUC	Cutoff	SN	SP	White et al. (2020)	
TMT-A raw	.64	-	-	-	.65*	≥62; .31/.90
TMT-A T-score	.68*	≤28	.36	.91	.68**	≤31; .27/.90
		≤30	**.44**	**.91**		
		≤33	.48	.86		
TMT-B raw	.60	-	-	-	.65*	≥285; .27/.84
TMT-B T-score	.66*	≤25	.20	.94	.67*	≤21; .19/.90
		≤26	**.28**	**.94**		
		≤28	.32	.86		

AACN 2025

25

Moving Forward: Validating More Robust& Broadly Applicable Embedded PVTs

1. Ensure representation of unimpaired and impaired populations in research and clinical samples used to validate embedded PVTs and demonstrate equivalency among those with impairment.
   - Subdivide and examine impaired subset of the sample separately

- Stroop Word Reading T-score (White et al., 2020)
  - Overall sample (valid vs. Invalid): ≤21T (54% SN / 91% SP)
  - Valid-Cognitively Impaired vs. Invalid: ≤21T (54% SN / 89% SP)
- Stroop Color Naming T-score (White et al., 2020)
  - Overall sample (valid vs. Invalid): ≤23T (46% SN / 90% SP)
  - Valid-Cognitively Impaired vs. Invalid: ≤20T (42% SN / 94% SP)

AACN 2025

26

Moving Forward: Validating More Robust& Broadly Applicable Embedded PVTs

2. Avoid embedded PVTs in which a single score is supposed to simultaneously convey validity and impairment status
   - How do you differentiate true impairment from invalidity with the same score?
   - How are normatively intact scores also supposed to indicate invalidity?
     - *Invalid Before Impaired Paradox* (Critchfield & Lichtenstein, 2017)

AACN 2025

27

Peak et al. (2024)-Unimpaired Subsample

TMT score	AUC	Cutoff	SN	SP	White et al. (2020)	
TMT-A raw	.80***	≥47	.60	.87	.88***	≥39; .58/.94
		≥49	**.56**	**.90**		
		≥53	.48	.92		
TMT-A T-score	.81***	≤37	.52	.95	.80***	≤40; .58/.90
		≤40	**.60**	**.90**		
		≤41	.64	.87		
TMT-B raw	.86***	≥100	.72	.85	.94***	≥80; .85/.90
		≥109	**.72**	**.90**		
		≥119	.68	.92		
TMT-B T-score	.88***	≤38	.64	.95	.87***	≤40; .62/.90
		≤40	**.72**	**.90**		
		≤42	.80	.85		

AACN 2025

28

Moving Forward: Validating More Robust& Broadly Applicable Embedded PVTs

2. Avoid embedded PVTs in which a single score is supposed to simultaneously convey validity and impairment status

- If using single scores, use scores robust against effects of impairment
- Combinatory/Algorithmic embedded PVTs often are more robust
  - RAVLT and RCFT Effort Scores (Boone et al., 2005; Lue et al., 2003)
  - RDS and its derivatives (Greiffenstein et al., 1994; Webber & Soble, 2018)
  - HVLT-R/BVMT-RD (Bailey et al., 2018)
  - Combining time and accuracy metrics (e.g., HVLT FC & completion time) (Cutter et al. (2022)
- Forced choice and other tried & true paradigms can be useful
  - CVLT-II/3 FC; RAVLT FC (Poreh et al., 2016; Soble et al. 2021)

AACN 2025

29

Moving Forward: Validating More Robust& Broadly Applicable Embedded PVTs

3. Develop different cut-scores for populations likely to present with no to minimal cognitive impairment vs. genuine cognitive impairment

- Not an entirely new concept:
  - Word Memory Test-Genuine Memory Impairment Profile (Green et al., 2011; Alverson et al., 2019)
  - Dot Counting Test (Boone et al., 2002; Hansen et al., 2023)
- Two potential pitfalls
  1. Full clinical history is not always known
  2. Even if an examinee is being evaluated for a condition like nonacute mTBI, it is not guaranteed that is the only condition at play (whether known or unknown)

AACN 2025

30

Moving Forward: Validating More Robust& Broadly Applicable Embedded PVTs

- **In sum, not every test can (or should) be made into a PVT!**
- Embedded PVTs (particularly single-score PVTs from tests known to be sensitive to genuine impairment) poorly validated in circumscribed, unimpaired populations risks artificially skewed psychometrics, poor replicability, and low generalizability to actual clinical populations
- Can become a problematic self-perpetuating cycle when these same embedded PVTs are then used as independent criterion measures to cross-validate other similarly problematic PVTs in future studies

AACN 2025

31

Future Directions: Moving Beyond Individual Embedded PVTs to Enhance Invalidity Detection

- Machine Learning
- Intraindividual Variability (Dispersion) Metrics

AACN 2025

32

Future Directions: Using Machine Learning to Enhance Invalidity Detection

- Neuropsychology landscape is evolving
  - Tele-neuropsychology
  - Greater emphasis on utility of "Big Data" (NNN)
- Potential to develop advanced statistical algorithms that can learn from existing data to enhance our ability to detect invalidity in a more complex and nuanced way
- *Using machine learning to detect noncredible test performance* (Finley et al., 2024)
  - ML identified PVT scores along with aspects of medical history, referral reason and disability status as most influential in differentiating valid from invalid performers
- *Detecting noncredible symptomology in ADHD evaluations using machine learning* (Finley et al., 2025)
  - ML can effectively identify noncredible symptom reporting using scores from multiple SVTs without predetermined cutoffs and 2 SVT elevations to best identified noncredible symptom reporting.

AACN 2025

33

Future Directions: Using Intraindividual Variability (IIV) Metrics to Enhance Invalidity Detection

- High IIV-dispersion has been investigated as a potential indicator of underlying CNS dysfunction/neuropsychological impairment (Costa et al., 2019; Jones et al., 2020)
  - Also posited as a potential indicator of invalid test engagement (Hill et al., 2013)
- *Does IIV index cognitive dysfunction, invalidity, or both among veterans* (Webber et al. 2024)
  - IIV-dispersion may be sensitive to both neurocognitive disorders and compromised engagement. Test validity (e.g. PVTs, behavioral signs of engagement) should be considered as an alternate explanation prior to interpretation of IIV as an indicator of cognitive impairment.
  - Currently being replicated among a larger, more diverse sample of mixed neuropsychiatric outpatients.
- *Dispersion and consistency based IIV in adult ADHD evaluations* (under review)
  - Results validated the use of IIV metrics, especially those based on dispersion, for differentiating invalid performance from genuine attentional impairments among adults undergoing ADHD evaluation, with higher levels of variability found among those with invalid performance.

AACN 2025

34

References

Alverson, W. A., O'Rourke, J. J. F., & Soble, J. R. (2019). The Word Memory Test genuine memory impairment profile discriminates genuine memory impairment from invalid performance in a mixed clinical sample with cognitive impairment. *The Clinical Neuropsychologist, 33*(8), 1428–1435.

Bailey, K. C., Soble, J. R., Bain, K. M., & Fullen, C. (2018). Embedded Performance Validity Tests in the Hopkins Verbal Learning Test-Revised and the Brief Visuospatial Memory Test-Revised: A Replication Study. *Archives of Clinical Neuropsychology, 33*(7), 897–908.

Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford.

Boone, K. B. (2021). *Assessment of feigned cognitive impairment: A neuropsychological perspective* (2nd ed.). Guilford Press.

Boone, K. B., Lu, P., & Herzberg, D. S. (2002). *The Dot Counting Test manual*. Western Psychological Services.

Boone, K. B., Lu, P., & Wen, J. (2005). Comparison of various RAVLT scores in the detection of noncredible memory performance. *Archives of Clinical Neuropsychology, 20*, 301–319.

Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology, 28*(7), 633–639.

Costa, A. S., Dogan, I., Schulz, J. B., & Reetz, K. (2019). Going beyond the mean: Intraindividual variability of cognitive performance in prodromal and early neurodegenerative disorders. *The Clinical Neuropsychologist, 33*(2), 369–389.

Cutler, L., Abeare, C. A., Messa, I., Holcomb, M., & Erdodi, L. A. (2022). This will only take a minute: Time cutoffs are superior to accuracy cutoffs on the forced choice recognition trial of the Hopkins Verbal Learning Test – revised. *Applied Neuropsychology: Adult, 29*(6), 1423–1439.

Davis, J. J., & Millis, S. R. (2014). Examination of performance validity test failure in relation to number of tests administered. The Clinical Neuropsychologist, 28(2), 199–214.

Dolorfo, R., Kivisto, A. J., Swier-Vosnos, A., & Elwood, L. (2023). Optimal per test cutoff scores and combinations of failure on multiple embedded performance validity tests in detecting performance invalidity in a mixed clinical sample. Applied Neuropsychology Adult, 30(6), 716–726.

Erdodi, L. A., & Lichtenstein, J. D. (2017). Invalid before impaired: an emerging paradox of embedded validity indicators. The Clinical Neuropsychologist, 31(6-7), 1029–1046.

Finley, J. A., Phillips, M. S., Soble, J. R., & Rodriguez, V. J. (2024). Detecting noncredible symptomology in ADHD evaluations using machine learning. Journal of Clinical and Experimental Neuropsychology, 46(10), 1015–1025.

AACN 2025

35

References

Finley, J. A., Robinson, A. D., Soble, J. R., & Rodriguez, V. J. (2024). Using machine learning to detect noncredible cognitive test performance. The Clinical Neuropsychologist, 1–19. Advance online publication. https://doi.org/10.1080/13854046.2024.2440885

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the word memory test and medical symptom validity test in groups with severe verbal memory impairment. Applied Neuropsychology, 18(2), 86–94.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. Psychological Assessment, 6(3), 218.

Hanson, N. D., Rhoads, T., Jennette, K. J., Reynolds, T. P., Ovsiew, G. P., Resch, Z. J., Critchfield, E. A., Marcotte, J. C., O'Rourke, J. J. F., & Soble, J. R. (2023). Validation of alternative dot counting test E-score cutoffs based on degree of cognitive impairment in veteran and civilian clinical samples. The Clinical Neuropsychologist, 37(2), 402–417.

Hill, B. D., Rohling, M. L., Boettcher, A. C., & Meyers, J. E. (2013). Cognitive intra-individual variability has a positive association with traumatic brain injury severity and suboptimal effort. Archives of Clinical Neuropsychology, 28(7), 640–648.

Jones, J. D., Burroughs, M., Apodaca, M., & Busch, J. (2020). Greater intraindividual variability in neuropsychological performance predicts cognitive impairment in de novo Parkinson's disease. Neuropsychology, 34(1), 24.

Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. Archives of Clinical Neuropsychology, 29(4), 364–373.

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. The Clinical Neuropsychologist, 23(5), 841–849.

Larrabee, G. J., Rohling, M. L., & Meyers, J. E. (2019). Use of multiple performance and symptom validity measures: Determining the optimal per test cutoff for determination of invalidity, analysis of skew, and inter-test correlations in valid and invalid performance groups. The Clinical Neuropsychologist, 33(8), 1354–1372.

Lezak, M. I., Finley, J. C. A., Barrette, K. S., VanLandingham, H. B., Pierczore, J., Bianco, J. M., Phillips, M. S., Coray, B. M., Schroeder, R. W., & Soble, J. R. (2025). Analysis of skew, examination of intercorrelations, and determining the optimal threshold for performance invalidity when 10 performance validity tests are administered during a neuropsychological evaluation. Journal of Clinical and Experimental Neuropsychology, 24(10), 989–1000.

Lu, P. H., Boone, K. B., Cozolino, L., & Mitchell, C. (2003). Effectiveness of the Rey-Osterrieth Complex Figure Test and the Meyers and Meyers recognition trial in the detection of suspect effort. The Clinical neuropsychologist, 17(3), 426–440.

AACN 2025

36

References

Martin, P. K., & Schroeder, R. W. (2020). Base rates of invalid test performance across clinical non-forensic contexts and settings. Archives of Clinical Neuropsychology, 35(6), 717–725.

Martin, P. K., Schroeder, R. W., & Olsen, D. H. (2022). Performance validity in the dementia clinic: Specificity of validity tests when used individually and in aggregate

across levels of cognitive impairment severity. The Clinical Neuropsychologist, 36(1), 165–188.

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. The Clinical Neuropsychologist, 29(6), 741–776.

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2024). Neuropsychologists' validity testing beliefs and practices: A survey of North American neuropsychologists and validity assessment experts. Archives of Clinical Neuropsychology. Online ahead of print.

Ovsiew, G. P., Resch, Z. J, Nayar, K., Williams, C. P., & Soble, J. R. (2020). Not so fast! Limitations of processing speed and working memory indices as embedded performance validity tests in a mixed neuropsychiatric sample. Journal of Clinical and Experimental Neuropsychology, 42(5), 472–484.

Pauk, A. M., Marcopulos, J. C., Chicote-Carroll, C., & Soble, J. R. (2024). Cross-validation of the Trail Making Test as a non-memory-based embedded performance validity test among veterans with and without cognitive impairment. Journal of Clinical and Experimental Neuropsychology, 46(1), 16–24.

Pearson N. C. S. (2009). Advanced clinical solutions for use with WAIS-IV and WMS-IV. Pearson Education.

Ponsh, A., Bezdicek, O., Kostrhova, I., Levin, J. B., & Dines, P. (2016). The Rey Auditory Verbal Learning Test forced-choice recognition task: Base-rate data and norms. Applied neuropsychology. Adult, 23(3), 155–161.

Schroeder, R. W., & Martin, P. K. (Eds.). (2021). Validity assessment in clinical neuropsychology: Evaluating and managing noncredible performance. Guilford.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. Archives of Clinical Neuropsychology, 22(2), 213–223.

Soble, J. R., Alvarez, W. A., Phillips, J. I., Critchfield, E. A., Fullen, C., O'Rourke, J. J. F., ... Marcopulos, J. C. (2020). Strength in numbers or quality over quantity? Examining the importance of criterion measure selection to define validity groups in performance validity test (PVT) research. Psychological Injury and Law, 13(1), 44–54.

Soble, J. R., Ovsiew, G. P., & Phillips, M. S. (2025). Performance validity tests in neurocognitive/neuropsychological assessments. In G. Young, T. Bailey, L. Giromini, J. R. Soble, R. Rogers, & B. E. Kapitt (Eds.), Handbook of psychological injury and law (pp. 275–294). Cham, Switzerland: Springer Nature.

37

References

Soble, J. R., Sharp, D. W., Carter, D. A., Jennette, K. J., Resch, Z. J., Ovsiew, G. P., & Critchfield, E. A. (2021). Cross-validation of a forced-choice validity indicator to enhance the clinical utility of the Rey Auditory Verbal Learning Test. Psychological Assessment, 33(6), 568–573.

Soble, J. R., Webber, T. A., & Bailey, K. C. (2021). An overview of common stand-alone and embedded PVTs for the practicing clinician: Cutoffs, classification accuracy, and administration times. In R. W. Schroeder & P. K. Martin (Eds.), Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance (pp. 126–149). New York: Guilford.

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. The Clinical Neuropsychologist, 23(2), 297–313.

Webber, T. A., Lotkiewicz, S., Woods, S. P., Miller, B., & Soble, J. R. (2024). Does neuropsychological intraindividual variability index cognitive dysfunction, an invalid presentation, or both? Preliminary findings from a mixed clinical older adult veteran sample. Journal of Clinical and Experimental Neuropsychology, 46(6), 535–556.

Webber, T. A., & Soble, J. R. (2018). Utility of various WAIS-IV Digit Span indices for identifying noncredible performance validity among cognitively impaired and unimpaired examinees. The Clinical neuropsychologist, 32(4), 657–670.

Wilde, D. J., Kerrigan, D., Barnerias, M. T., Ovsiew, G. P., Resch, Z. J., & Soble, J. R. (2020). Cross-validation of non-memory-based embedded performance validity tests for detecting invalid performance among patients with and without neurocognitive impairment. Journal of clinical and experimental neuropsychology, 42(5), 459–472.

Young, J. C., Roper, B. L., & Arentsen, T. J. (2016). Validity testing and neuropsychology practice in the VA healthcare system: results from recent practitioner survey. The Clinical Neuropsychologist, 30(4), 497–514.

38

Domain Specificity of Over-Report Symptom Validity Scales on the PAI & MMPI

Patrick Armistead-Jehle, PhD, ABPP-CN
Munson Army Health Center
Fort Leavenworth, KS

AACN 2025

1

Disclaimer/Disclosures

- Any opinions, findings, conclusions or recommendations expressed are those of the presenter and do not necessarily reflect the views of the U.S. Government, Department of Defense, or Defense Health Agency, and no official endorsement should be inferred.
- No financial relationships or other conflicts of interest to disclose.

AACN 2025

2

Over-Report Domain Specificity

- Conceptual distinction between domains of over-report response bias
  1. Psychological
  2. Cognition
  3. Somatic
- Based on assumption that respondents will uniquely overreport specific symptom sets

AACN 2025

3

Over-Report Domain Specificity

- SVT over-report scales rarely elevate in isolation and examinees often fail to distinguish between symptom sets
- Co-elevation tends to be a rule, rather than the exception
- Thus, the distinction between cognitive, psychological, and somatic domains maybe more conceptual and less likely to appear in applied settings.

AACN 2025

4

PAI Over-Report Domains & Scales

Psychological

The OGs:
- NIM
- RDF (Supplemental)
- MAL (Supplemental)

New(ish):
- Multiscale Feigning Index (MFI)
- Hong Malingering Index (HongM)
- Negative Distortion Scale (NDS)

Cognitive
- Cognitive Bias Scale (CBS)
- CBS SOS 1 – 3
- Feigned Adult ADHD Index (FAA)*

Somatic
- Malingered Pain-Related Disability Discriminant Function (MPRD)*

* Very few studies

AACN 2025

5

PAI Over-Report Scale Construction

- NIM: 9 items rarely endorsed by clinical and normative samples (represent bizarre or unlikely symptoms or present an exaggerated/distorted impression of self/circumstances)
- MAL: 8 configural features of the PAI profile that were more often seen in those simulating mental disorder than clinical patients
- RDF: Regression-based scale using beta weights across 20 different scales and subscales that distinguished simulators from patients

AACN 2025

6

PAI Over-Report Scale Construction

- MFI: Average of multiple clinical scales predictive of SIRS failures.
- HongM: Weighted average of five scales selected via stepwise discriminant function analysis
- NDS: 15 infrequently endorsed items selected from clinical scales among a sample of involuntarily committed psychiatric patients.

PAI Over-Report Scale Review:

Kurtz, J.E. & McCredie, M.N. (2022). Exaggeration or Fabrication? Assessment of Negative Response Distortion and Malingering with the Personality assessment Inventory. Psychological Injury and Law, 15, 37-47.

AACN 2025

7

PAI Over-Report Scale Construction

- CBS: 10 PAI items found to predict PVT failure
- CBS SOS 1-3: Scale level content found to predict PVT failure

Scale	Content
CBS-SoS1	Avg of NIM, SOM, DEP, ANX, SCZ, & SUI
CBS-SoS2	Logistically derived from NIM, SOM, DEP, ANX, SCZ and SUI
CBS-SoS3	Avg of SOM-C, DEP, SOM-S, SCZ, NIM, and PAR-R.

AACN 2025

8

PAI Over-Report Scale Construction

- FAA Index
  - Scale level: Regression-based algorithm comprised of PIM, SCZ-T, ANT-S, and DEP-C that differentiated genuine ADHD vs ADHD simulators.
    - Then applied to other diagnostic groups (mood/anxiety, controls, etc.).
  - Item level: Regression-based algorithm comprised of 24 items that differentiated between feigned ADHD and valid performers.
- Malingered Pain-Related Disability Discriminant Function (MPRD)
  - Regression based algorithm consisting of 32 scales and subscales found to distinguish actual chronic pain patients from malingering pain disability simulators.

AACN 2025

9

PAI Over-Report Scale Correlations: Original Normative Sample

	NIM	MAL	RDF	NDS	HMI	MFI
NIM	-	.49	.38	.68	.83	.74
MAL	.62	-	.26	.43	.50	.47
RDF	.10	.11	-	.41	.44	.40
NDS	.75	.57	.19	-	.71	.72
HMI	.88	.61	.18	.78	-	.89
MFI	.80	.59	.14	.80	.89	-

Above diagonal: Community adult normative sample (n=1000)
Below diagonal: Clinical standardization sample (n=1246)

From Kurtz & McCredie (2022)

AACN 2025

10

PAI Scale Correlations: Other Studies

Schroeder, Bieu, & Snodgrass (2025)

	NIM	MAL	RDF	MFI	HMI	CBS	SOS-1	SOS-2	SOS-3
NIM	-	.60	.23	.73	.83.	.66	.78	.76	.78
MAL	.52	-	.19	.61	.61	.49	.58	.54	.57
RDF	-.09	-.09	-	.27	.27	.29	.33	.25	.32
MFI	.75	.69	.06	-	.88	.77	.95	.88	.93
HMI	.83	.55	.05	.84	-	.67	.83	.78	.82
CBS	.65	.46	-.10	.67	.52	-	.80	.84	.82
SOS-1	.75	.61	.07	.85	.71	.77	-	.89	.94
SOS-2	.71	.47	-.13	.79	.63	.81	.82	-	.95
SOS-3	.80	.53	-.07	.87	.74	.76	.86	.92	-

Above diagonal PVT pass (n=184)
Below diagonal PVT fail (n=51)

AACN 2025

11

PAI Over-Report Scale Correlations: Other Studies

Schroeder, Bieu, & Snodgrass (2025)

- "Most of the over-reporting validity scales had strong to very strong correlations with each other regardless of whether they were designed to assess for over-reporting psychiatric or cognitive issues."
- " This suggests that an underlying construct inherent throughout these validity scales is that of over-reporting, regardless of domain…"

AACN 2025

12

PAI Over-Report Scale Correlations: Other Studies

Schroeder, Bieu, & Snodgrass (2025)

Scale/Cut	AUC	Sen	Spec
NIM ≥ 81	.63	.18	.93
MAL ≥ 3	.59	.10	.95
RDF ≥ 66	.47	.04	.92
MFI ≥ 78	.70	.14	.92
HMI ≥ .72	.64	.20	.90
CBS ≥ 18	.71	.29	.91
CB-SOS1≥ 78	.67	.16	.91
CB-SOS2 ≥ 5.4	.70	.28	.91
CB-SOS3 ≥ 74	.71	.28	.90

AACN 2025

13

PAI Over-Report Scale Correlations: Other Studies

Schroeder, Bieu, & Snodgrass (2025)

"While an underlying construct related to over-report appears common to most of the over-report validity scales, the CBS and two of the CB-SOS…better identified noncredible memory impairment than the other over-report validity scales, indicating that there is utility in using these scales within neuropsychological samples"

AACN 2025

14

PAI Over-Report Scale Correlations: Other Studies

Herring, Albertorio, Diehl, & Ingram (Submitted)

Meta-Analysis of the PAI over-reporting and supplemental indicators

General Random Effects for PAI Over-Reporting Scales

Scale	k	g
NIM	67	1.56
MAL	52	1.11
RDF	48	1.03
NDS	16	1.81
MFI	7	1.87
HMI	6	1.89
CBS	7	0.87
SOS-1	5	0.94
SOS-2	4	0.87
SOS-3	5	1.00

AACN 2025

15

PAI Over-report Scale Correlations: Other Studies

Herring, Albertorio, Diehl, & Ingram (Submitted)

Meta-Analysis of the PAI Over-Reporting and Supplemental Indicators.

"…the PAI scales in our study appear to perform generally equitably to one another…suggesting that the scales capture a broad trait (e.g., response bias) rather than specific domains or approaches."

AACN 2025

16

PAI Scale Correlations: Other Studies

Shura, Ingram, Schroeder, & Armistead-Jehle (Submitted). Interpreting the Personality Assessment Inventory (PAI) Validity Scales: Leveraging Population-level Veteran Affairs (VA) data from 2008 to 2024.

- VA Corporate Data Warehouse with all PAIs administered within the VA from 2008-2024
  - N= 36,830

AACN 2025

17

PAI Over-Report Scale Correlations

Shura, Ingram, Schroeder, & Armistead-Jehle (Submitted)

	NIM	MAL	RDF	MFI	CBS	SOS1	SOS2	SOS3	HongM	NDS
NIM	-									
MAL	.68	-								
RDF	.22	.21	-							
MFI	.77	.63	.25	-						
CBS	.70	.54	.29	.86	-					
SOS1	.85	.65	.30	.93	.86	-				
SOS2	.80	.59	.22	.91	.88	.91	-			
SOS3	.83	.63	.30	.94	.88	.95	.96	-		
HongM	.88	.65	.29	.88	.73	.86	.81	.86	-	
NDS	.72	.59	.19	.73	.65	.78	.73	.75	.72	-

AACN 2025

18

PAI Over-Report Scale Correlations: CBS and CB-SOS papers

Several done…

Armistead-Jehle, P. & Ingram, P.B. (2023)

Armistead-Jehle, P. Ingram, P.B., & Morris, N.M. (2020)

Boress, K., Gaasedelen, O.J, Croghan, A., Johnson, M.K., Caraher, K., Basso, M.R., & Whiteside, D.M. (2022a)

Boress, K., Gaasedelen, O.J., Croghan, A., Johnson, M.K., Caraher, K., Basso, M.R., & Whiteside, D.M. (2022b)

Gaasedelen, O. J., Whiteside, D. M., Altmaier, E.,Welch, C., & Basso, M. R. (2019)

Ingram, P.B., Armistead-Jehle, P., Morris, N.M., & Herring, T.T. (2023)

Shura, R. D., Ingram, P. B., Miskey, H. M., Martindale, S. L., Rowland, J. A., & Armistead-Jehle, P. (2023).

…but associations among other non-cognitive SVTs not reported

AACN 2025

19

PAI Over-Report Scales: Are they domain specific?

AACN 2025

20

Over-Report Domains/Scales – MMPI-2-RF/3

Psychological	Cognitive	Somatic
• Infrequent Responses (F-r/F)	• Response Bias Scale (RBS)	• Symptom Validity Scale (FBS-r/FBS)
• Infrequency – Psychopathology (Fp-r/Fp)		• Infrequent Somatic Responses Scale (Fs)

AACN 2025

21

MMPI-2-RF Over-Report Scale Construction

- F-r: 32 items endorsed by ≤ 10% of normative sample
- Fp-r: 21 items predictive of symptom over-report in individuals tested in settings with high base rates of severe psychopathology.
  - No more than 20% of individuals with psychopathology had responded to the item in the keyed direction.
- Fs: 16 items endorsed by 25% or less in several large samples of medical patients

AACN 2025

22

MMPI-3 Over-Report Scale Construction

- F: 35 items rarely answered in the keyed direction by members of the MMPI-3 normative sample
  - ≤ 15% of normative sample (10% too stringent to produce enough items)
- Fp: 21 items rarely answered in the keyed direction by individuals with genuine, severe psychopathology.
  - MH sample used to identify Fp-r items that did not meet 20% criterion and replace them with MMPI-3 items that did.
- Fs: 16 items endorsed by 25% or less in several large samples of medical patients
  - Medical sample used to identify items that did not meet 25% criterion and replace them with MMPI-3 items that did.

AACN 2025

23

MMPI-2-RF/3 Over-Report Scale Construction

- RBS: 28 items identified as predictive of PVT failure (cognitive symptoms).
- FBS-r/FBS: 30 items meant to identify non-credible responding in test takers involved in personal injury (civil) litigation (over-report of cognitive and somatic symptoms)

AACN 2025

24

MMPI-2-RF Over-Report Scale Correlations: Manual

	F-r	Fp-r	Fs	FBS-r	RBS
F-r	-	.48	.52	.35	.47
Fp-r	.48	-	.34	.18	.33
Fs	.49	.30	-	.30	.41
FBS-r	.43	.23	.38	-	.39
RBS	.54	.38	.41	.47	-

Above Diagonal: Normative Sample - Men (N=1138).

Below Diagonal: Normative Sample – Women (N=1138)

	F-r	Fp-r	Fs	FBS-r	RBS
F-r	-	.61	.68	.66	.77
Fp-r	.61	-	.45	.39	.51
Fs	.65	.43	-	.59	.65
FBS-r	.64	.36	.66	-	.69
RBS	.79	.53	.61	.71	-

Above Diagonal: Outpatient Sample - Men (N=410).

Below Diagonal: Outpatient Sample – Women (N=610)

AACN 2025

25

MMPI-3 Over-Report Scale Correlations: Manual

	F	Fp	Fs	FBS	RBS
F	-	.49	.51	.43	.53
Fp	.51	-	.32	.17	.32
Fs	.52	.35	-	.46	.43
FBS	.48	.23	.51	-	.49
RBS	.57	.36	.50	.57	-

Above Diagonal: Normative Sample - Men (N=810).

Below Diagonal: Normative Sample – Women (N=810)

	F	Fp	Fs	FBS	RBS
F	-	.62	.60	.46	.69
Fp	60	-	.44	.22	.39
Fs	.59	.43	-	.62	.65
FBS	.52	.32	.63	-	.61
RBS	.67	.48	.65	.60	-

Above Diagonal: Outpatient Sample - Men (N=339).

Below Diagonal: Outpatient Sample – Women (N=452)

AACN 2025

26

MMPI-3 Overreport Scales: Morris et al

Morris et al (2021). Evaluating the performance of the MMPI-3 over-reporting scales: Sophisticated simulator and the effect of comorbid conditions. TCN

310 undergrads assigned to one of three coached conditions (PTSD, mTBI, or both) or honest response.

AACN 2025

27

MMPI-3 Overreport Scales: Morris et al

Table 2. ANOVA results.

		1. Control		2. PTSD		3. mTBI		4. PTSD+mTBI	
		n=85		n=81		n=70		n=74	
	F	M (SD)	% ≥ CS	M (SD)	% ≥ CS	M (SD)	% ≥ CS	M (SD)	% ≥ CS
F	33.54*	53.5 (15.3)[a]	1.20%	89.2 (32.3)	49.40%	79.8 (27.1)	37.10%	91.2 (31.1)	50.0%
Fp	28.55*	56.4 (15.4)[a]	1.20%	89.5 (31.8)	48.10%	79.1 (28.8)	37.10%	91.8 (30.9)	51.4%
Fs	36.99*	54.3 (13.2)[a]	1.20%	87.1 (28.8)	44.40%	84.1 (27.9)	30.00%	88.4 (27.9)	43.2%
FBS	33.85*	52.9 (10.6)[a]	0.00%	69.5 (15.1)	0.00%	71.7 (13.7)	0.00%	70.6 (15.4)	0.0%
RBS	36.31*	53.1 (11.7)[a]	1.20%	82.0 (25.6)	32.10%	77.6 (20.1)	17.10%	82.8 (24.6)	31.1%
Cohen's d Effect Size									
	1 v 2	1 v 3	1 v 4	2 v 3	2 v 4	3 v 4			
F	**1.42**	**1.22**	**2.47**	−0.32	0.06	−0.39			
Fp	**1.33**	**1.00**	**1.47**	−0.34	0.07	−0.42			
Fs	**1.47**	**1.57**	**1.59**	−0.11	0.05	−0.17			
FBS	**1.28**	**1.55**	**1.35**	0.15	0.07	−0.07			
RBS	**1.46**	**1.52**	**1.57**	−0.19	0.03	0.23			

AACN 2025

28

MMPI-3 Over-Report Scale Correlations

Tylicki, Gervais, & Ben-Porath (2022). Examination of the MMPI-3 Over-reporting sales in a forensic disability sample. TCN

550 non-head injury disability related referrals administered a range of PVTs and SVTs with MMPI-3

	MMPI-3				
	F	Fp	Fs	FBS	RBS
PAI NIM	.75	.58	.60	.43	.60
PAI MAL	.50	.44	.39	.30	.40

AACN 2025

29

MMPI-2-RF Over-Report Scale Correlations

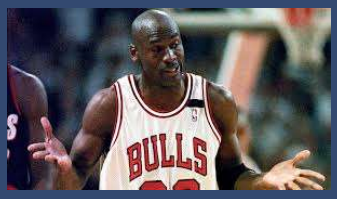
- VA CDW Data: N = 195,837
- Unpublished

	F-r	Fp-r	Fs	FBS	RBS
F-r	-				
Fp-r	.70	-			
Fs	.73	.60	-		
FBS-r	.65	.46	.66	-	
RBS	.81	.61	.70	.73	-

AACN 2025

30

MMPI-2-RF/3 Over-Report Scales: Are they domain specific?



AACN 2025

31

Multivariate Base Rate Approaches

- Presume over-report is generalized and is a single construct…

…then examine pattern across scales versus individual items

- MBR: Method of over-report detection relying on multiple scale scores
  - Infrequent scale elevation combinations rather than a combination of infrequently endorsed single items.
- Precedent with PAI (e.g., MFI and CB-SOSs)

AACN 2025

32

Multivariate Base Rate Approaches: MMPI

Ingram et al (2024). Development and Initial validation of Scale of Scales (SOS) over reporting scores for the MMPI family of instruments. JCEN

Two Methods:

1. SOS Mean Score
   - Average RC scale scores (MMPI-3 = 7; MMPI-2-RF = 8)
2. SOS Elevation Frequency
   - Number of selected substantive scales ≥ 65T

Two Samples:

1. Simulation Design
   - College students (n= 318) assigned to feigning or honest conditions
2. Known-group Design
   - AD SMs (n=290) with failed PVTs as criterion.

AACN 2025

33

Multivariate Base Rate Approaches: MMPI

Simulation Design

SOS Mean	Score	g	Sens	Spec
MMPI-2-RF	70	1.08	.52	.89
MMPI-3	70	1.01	.47	.92
SOS Elevation Frequency				
MMPI-2-RF	6	.84	.39	.92
MMPI-3	5	.80	.37	.92

Known Groups Design

SOS Mean	Score	g	Sens	Spec
MMPI-2-RF	70	.95	.18	.94
SOS Elevation Frequency				
MMPI-2-RF	6	.67	.21	.93

Incremental Validity (Logistic Regression)

- Both SOS scales improved prediction over F, Fs, and FBS.
- Neither SOS scale improved over Fp-r
- Mixed findings for RBS

AACN 2025

34

Multivariate Base Rate Approaches: MMPI

Primary Findings:

1. MBR offers a novel and potentially useful addition to existing MMPI over-report detection methods
2. Classification accuracy similar to existing over-report scales
3. Both SOS Mean and Elevation Frequency approaches provide similar classification performance, suggesting a robustness to the MBR approach.

AACN 2025

35

Multivariate Base Rate Approaches: PAI

Aita et al (submitted) Journal of Personality Assessment

Study Aims:

1. Describe MBR of elevated scores across three samples
2. Determine whether MBR of elevated PAI scores can identify non-credible PAI profiles among PTSD simulators while differentiating them from those with genuine PTSD.
3. Compare diagnostic effectiveness of MBR to existing over-report SVTs

AACN 2025

36

Multivariate Base Rate Approaches: PAI

Three Samples:

1. PTSD (Veterans = 111; SMs = 141)
2. Mixed Mood Disorders (Veterans = 32; SMs = 361; Civilian Adults = 131)
3. PTSD Simulators (n=160)

AACN 2025

37

Multivariate Base Rate Approaches: PAI

Comparison	PAI Variable	AUC	Cutoff (≥j)	Sn	Sp	Accuracy
PTSD vs. PTSD sim	Scales Standard j	.86	8	.78	.92	86.7%
	Scales Skyline j	.80	2	.65	.83	76.7%
	Subscales Standard j	.84	18	.72	.90	81.5%
	Subscales Skyline j	.81	4	.69	.84	77.3%
	NIM	.86	92	.65	.94	83.0%
	MFI	.82	77	.76	.77	77.9%

AACN 2025

38

Multivariate Base Rate Approaches: PAI

Comparison	PAI Variable	AUC	Cutoff (≥j)	Sn	Sp	Accuracy
Mood dx vs. PTSD sim	Scales Standard j	.92	8	.78	.98	93.9%
	Scales Skyline j	.87	1	.78	.90	87.3%
	Subscales Standard j	.90	14	.81	.96	91.2%
	Subscales Skyline j	.88	3	.74	.92	87.1%
	NIM	.91	92	.65	.99	90.8%
	MFI	.91	77	.76	.96	92.0%

AACN 2025

39

Multivariate Base Rate Approaches: PAI

Primary Findings:

- MBR cutoffs identified from receiver-operating characteristic curve analyses yielded robust sensitivity (.650-.806) and specificity (.833-.984) in differentiating genuine PTSD and mood disorder groups from PTSD simulators.
  - **Clearly needs to be replicated in clinical (non-simulation) samples**
- MBRs were useful in differentiating genuine from simulated psychopathology, consistent with broader scale-based infrequency approaches (e.g., NIM, MFI).

AACN 2025

40

Conclusions/Future Research

- Over-report scale domain distinctions may not be exceptionally well defined
  - SEM and Factor Analytic Studies
- MBR approaches hold promise for broadband measures.
  - MBR studies validation studies

AACN 2025

41

References

Aita, S.L. Montgomery, E.L., Caron, J.E., Pagano, L.A., Broggi, M., Ingram, P.B., Erickson, S.C., Borgagna, N.C. Moncreief, G.C., Roth, R.M., Calamia, M.R., Armistead-Jehle, P, & Hill, B.D. (submitted). Multivariate base rates of standard and skyline cut off elevations on the Personality Assessment Inventory: Do they distinguish simulated from genuine PTSD? Journal of Personality Assessment.

Herring, T. T., Albertorio, A.D., Diehl, K.J., & Ingram, P.B. (submitted). A meta-analysis of the Personality Assessment Inventory (PAI) over-reporting scales and supplemental indicators. Journal of Psychopathology and Behavioral Assessment.

Ingram, P.B., Keen, M.A., Greene, T.E., Morris, C., & Armistead-Jehle, P. (2024). Development and initial validation of the Scale of Scales (SoS) overreporting scores for the MMPI family of instruments. Journal of Clinical and Experimental Neuropsychology, 46(2), 95-110.

Kurtz, J.E. & McCredie, M.N. (2022). Exaggeration or Fabrication? Assessment of Negative Response Distortion and Malingering with the Personality assessment Inventory. Psychological Injury and Law, 15, 37-47.

Morris, N.M., Mattera, J., Golden, B., Moses, S. & Ingram, P.B. (2022). Evaluating the performance of the MMPI-3 over-report scales: Sophisticated simulators and the effect of comorbid conditions. The Clinical Neuropsychologist, 36, 2361-2369.

Schroeder, R.W., Bieu, R., K., & Snodgrass, M. (2025). Comparing the cognitive bias scale and the cognitive bias scale of scales to other personality assessment inventory validity scales for detecting noncredible memory dysfunction in a clinical veteran sample. JCEN. DOI: 10.1080/13803395.2025.2464635.

Shura, R.D. Ingram, P.B., Schroeder, R.W., & Armistead-Jehle, P. (submitted). Interpreting the Personality Assessment (PAI) validity scales: Leveraging population-level Veteran Affairs data from 2008 to 2024. Psychological Assessment.

Tylicki, J.L., Rai, J.K., Arends, P., Gervais, R.O., & Ben-Porath, Y.S. (2021). A comparison of the MMPI-2-RF and PAI overreporting indicators in a civil forensic sample with emphasis on the Response Bias Scale (RBS) and the Cognitive Bias Scale (CBS). Psychological Assessment, 33, 71-83.

AACN 2025

42

?

Patrick.J.Armistead-Jehle.civ@health.mil

AACN 2025

43