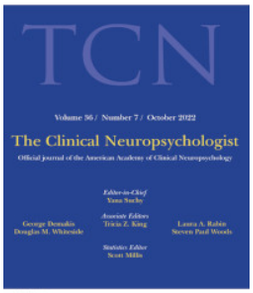

The Clinical Neuropsychologist

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ntcn20

Validation of the personality assessment inventory (PAI) cognitive bias (CBS) and cognitive bias scale of scales (CB-SOS) in a post-deployment veteran sample

Robert D. Shura, Paul B. Ingram, Holly M. Miskey, Sarah L. Martindale, Jared A. Rowland & Patrick Armistead-Jehle

To cite this article: Robert D. Shura, Paul B. Ingram, Holly M. Miskey, Sarah L. Martindale, Jared A. Rowland & Patrick Armistead-Jehle (2022): Validation of the personality assessment inventory (PAI) cognitive bias (CBS) and cognitive bias scale of scales (CB-SOS) in a post-deployment veteran sample, The Clinical Neuropsychologist, DOI: 10.1080/13854046.2022.2131630

To link to this article: https://doi.org/10.1080/13854046.2022.2131630

Published online: 22 Oct 2022.

THE CLINICAL NEUROPSYCHOLOGIST
https://doi.org/10.1080/13854046.2022.2131630

Validation of the personality assessment inventory (PAI) cognitive bias (CBS) and cognitive bias scale of scales (CB-SOS) in a post-deployment veteran sample

Robert D. Shura[a,b,c], Paul B. Ingram[d,e], Holly M. Miskey[a,b,c], Sarah L. Martindale[a,b,c], Jared A. Rowland[a,b,c] and Patrick Armistead-Jehle[f]

[a]W. G. (Bill) Hefner VA Healthcare System, Salisbury, NC, USA; [b]VA Mid-Atlantic (VISN 6) Mental Illness Research, Education, and Clinical Center (MIRECC), Durham, NC, USA; [c]Wake Forest School of Medicine, Winston-Salem, NC, USA; [d]Texas Tech University, Lubbock, TX, USA; [e]Dwight D. Eisenhower Veteran Affairs Medical Center, Eastern Kansas Veteran Healthcare System, Leavenworth, KS, USA; [f]Munson Army Health Center, Fort Leavenworth, KS, USA

ABSTRACT

Objective: The present study evaluated the function of four cognitive, symptom validity scales on the Personality Assessment Inventory (PAI), the Cognitive Bias Scale (CBS) and the Cognitive Bias Scale of Scales (CB-SOS) 1, 2, and 3 in a sample of Veterans who volunteered for a study of neurocognitive functioning. **Method:** 371 Veterans (88.1% male, 66.1% White) completed a battery including the Miller Forensic Assessment of Symptoms Test (M-FAST), the Word Memory Test (WMT), and the PAI. Independent samples *t*-tests compared mean differences on cognitive bias scales between valid and invalid groups on the M-FAST and WMT. Area under the curve (AUC), sensitivity, specificity, and hit rate across various scale point-estimates were used to evaluate classification accuracy of the CBS and CB-SOS scales. **Results:** Group differences were significant with moderate effect sizes for all cognitive bias scales between the WMT-classified groups (d = .52–.55), and large effect sizes between the M-FAST-classified groups (d = 1.27–1.45). AUC effect sizes were moderate across the WMT-classified groups (.650–.676) and large across M-FAST-classified groups (.816–.854). When specificity was set to .90, sensitivity was higher for M-FAST and the CBS performed the best (sensitivity = .42). **Conclusion:** The CBS and CB-SOS scales seem to better detect symptom invalidity than performance invalidity in Veterans using cutoff scores similar to those found in prior studies with non-Veterans.

ARTICLE HISTORY
Received 8 June 2022
Accepted 27 September 2022
Published online 22 October 2022

KEYWORDS
Personality assessment inventory; cognitive bias scale; CB-scale of scales; veterans

Assessment of response bias, including performance (PVT) and symptom validity tests (SVT), is a key component of a neuropsychological evaluation. Broadband personality measures include SVTs, which rely on detecting invalid responding based on self-report symptom levels and are typically focused on identifying over-reported psychological symptoms (Rogers, 2008). However, detecting invalid reporting of cognitive symptoms, which is related to but distinct from performance validity, is critical to

CONTACT Robert D. Shura Robert.Shura2@va.gov W. G. (Bill) Hefner VA Healthcare System, 1601 Brenner Ave (11-M), Salisbury, NC, 28144, USA.

neuropsychological evaluations because cognitive complaints frequently drive these referrals. Because cognitive complaints may not fully align with actual performance (Armistead-Jehle et al., 2012; Pearman, 2009) and most traditional SVTs focus on psychological complaints, SVTs that identify exaggerated, misrepresented, or otherwise invalid forms of cognitive symptom endorsement are critically needed within broad-band instruments to allow integrative interpretation of test validity (Sherman et al., 2020; Sweet et al., 2021).

The Personality Assessment Inventory (PAI; Morey, 2007) is a popular multi-scale, self-report instrument that includes several symptom validity scales in addition to its indices of psychopathology. The standard validity scales of the PAI, as well as several supplemental scales, have shown to be effective SVTs (Hawes & Boccaccini, 2009); however, the PAI has two weaknesses in the detection of overreporting. First, supplemental validity scales are rarely cross validated despite psychometric promise (McCredie & Morey, 2018). Lack of validation limits generalizability to distinct populations that require special considerations. Second, until recently, none of the validity scales on the PAI assessed cognitive symptom overreporting.

The Cognitive Bias Scale (CBS) was created using a civilian sample of mostly clinical and a small number of forensic participants (Gaasedelen et al., 2019). A variety of methods were used to ultimately identify 10 items. Although some of the items clearly focus on cognitive complaints, others focus on broader patterns of anxiousness and physical discomfort congruent with the non-cognitive/somatic scales from which some items were drawn. As a post-creation validation (Gaasedelen et al., 2019), CBS was contrasted across valid ($n = 257$) and invalid ($n = 49$) groups composed of those who failed two PVTs, including at least one stand-alone measure (Word Memory Test [WMT] or Test of Memory Malingering [TOMM]). Classification demonstrated large effects (AUC = .72), resulting in a recommended cutoff score ≥ 19 to produce desired levels of specificity (.96). Three subsequent cross-validations were published for the CBS. In outpatients drawn from an academic medical center, Boress et al. (2022) observed comparable performance at the ≥ 19 cutoff score (specificity = .92 and sensitivity = .35). In a civil forensic setting, Tylicki et al. (2021) contrasted CBS performance to a comparable cognitive SVT from which the CBS development was modeled (i.e. Response Bias Scale [RBS]; Gervais et al., 2007), observing similar classification statistics (.43 and .88 at the ≥ 19 cutoff score, respectively). However, these authors noted that achievement of a desired false positive rate ≤ 10% (specificity ≥ .90) required a slightly higher cutoff score of 20. When cross-validated in an active-duty service member sample (Armistead-Jehle et al., 2020), CBS achieved comparable performance at lower cut-scores using multiple invalid PVTs to group participants. A cutoff score of ≥ 14 produced a sensitivity of .32 and a specificity of .92. Therefore, in contrast to civilian samples, active-duty samples needed a lower cutoff score to achieve similar diagnostic accuracy.

Aside from the CBS, Boress et al. (2022) developed three additional SVTs targeting cognitive symptoms, the Cognitive Bias Scale of Scales (CB-SOS) 1, 2, and 3. In contrast to the CBS, which used individual item responses, CB-SOS 1, 2, and 3 have a scale-based construction: CB-SOS-1 is the average T score of six scales (NIM, SOM, DEP, ANX, SCZ, and SUI), CB-SOS-2 weighted those scales based on a binary logistic regression of PVT failure, and CB-SOS-3 is the average of the seven scales or subscales with the

highest AUC (SOM-C, DEP-P, SOM-S, ANX-P, SCZ, NIM, and PAR-R). Comparatively, CB-SOS-3 performed best with large effect classification accuracy (AUC = .75) and had equivalent sensitivity and specificity (.38 and .92 at the cutoff score of T ≥ 74, respectively) to the CBS scale (Boress et al., 2022; Tylicki et al., 2021). Scales and subscales selected for the various CB-SOS include broad measurement of some of the same item-level content measured by CBS (e.g. SOM-S content is included in the calculation of SOM, which is incorporated into CB-SOS-1 and a specific ANX-P item in CBS overlaps with the inclusion of ANX-P in CB-SOS-3). One independent study evaluated the CB-SOS and found medium to large effects between groups passing and failing PVTs, observing similar performance as the CBS with active-duty personnel (Ingram et al., in press). All of the prior studies focused on using PVTs as a criterion for the CBS and SOS, and none evaluated the new scales in comparison to other SVTs.

Current study

The present investigation is the first study to evaluate the CBS and CB-SOS scales in a sample of post-deployment Veterans. The creation of the CBS and CB-SOS scales helps address the absence of cognitive SVTs on the PAI; Table 1 presents a summary of sensitivity/specificity at tested cutoff scores for these scales. The scales have been minimally cross validated, which is rare for supplemental validity scales, even for a small number of studies (McCredie & Morey, 2018). However, active-duty military populations have consistently differed on CBS and CB-SOS relative to civilian samples in classification accuracy and recommended cutoff scores (i.e. Armistead-Jehle et al., 2020; Ingram et al., 2020). Although Veteran samples share common experiences with active-duty samples, Veteran populations are unique in their symptom validity assessment needs (Ray, 2017; Russo, 2013, 2018; Worthen & Moering, 2011) due to the disability system tied to such assessments (DeViva & Bloem, 2003; Freeman et al., 2008), potentially motivating Veterans to perform poorly on neuropsychological assessments. Additionally, the Department of Veterans Affairs and Department of Defense follow different missions and life circumstances for service members and Veterans are unique (for a discussion on this within the validity context, see Shura et al., 2021a). Because of this, elevated symptom validity concerns are frequent and differ by evaluation context (Ingram et al., 2020), requiring that SVTs assessing all domains of invalid content-based over-reporting are available within validated assessments. Given the medical and psychiatric complexity of Veteran samples and omnipresence of disability incentives in the VA, we predicted that 1) cutoff scores would need to be higher than observed civilian cut scores, similar to what has been observed with other disability samples (e.g. Tylicki et al., 2021) and consistent with the elevated scale scores typically observed in veterans on broadband measures generally (Ingram et al., 2020) and across PAI validity scale scores specifically (e.g. Ingram et al., 2022; Mozley et al., 2005), 2) effect size metrics evaluating classification and between group comparison would be lower and fall within the medium range, consistent with past work on the CBS and CB-SOS scales (Armistead-Jehle et al., 2020; Ingram et al., in press; Boress et al., 2021; Tylicki et al., 2021), and 3) our sensitivity would be low at adequate specificity, as is typically in research on broadband measure validity scale research (see Rogers, 2008).

Table 1. Summative cutoff scores recommended across existing CBS/CB-SOS literature.

Scale and ≥ cutoff	Gaasedelen et al. (2019)		Armistead-Jehle et al. (2020)		Tylicki et al. (2021)		Boress et al. (2022)		Ingram et al., in press	
	Sen.	Spec.	Sen.	Spec.	Sen.	Spec.	Sen.	Spec.	Sen.	Spec.
CBS										
14	0.47	0.83	0.38	0.82	0.85	0.58				
16	0.37	0.90	**0.27**	**0.93**	0.63	0.71			**0.22**	**0.93**
19	**0.31**	**0.96**	0.08	0.97	0.43	0.88			0.07	0.97
CB-SOS1										
75							0.38	0.86	**0.14**	**0.97**
78							**0.29**	**0.90**	0.07	0.99
80							0.18	0.96	0.02	0.99
CB-SOS2										
5.1							0.47	0.84	0.22	0.94
5.3							**0.41**	**0.90**	0.14	0.96
5.7							0.29	0.95	–	–
CB-SOS3										
73							0.38	0.89	0.16	0.97
74							**0.38**	**0.92**	0.12	0.97
77							0.35	0.96	0.09	0.98
Sample	Mixed neuro outpatient		Active-duty military		Disability claimants		Mixed neuro outpatient		Active-duty military	
N	306		197		588		332		288	

Note. CBS = Cognitive Bias Scale; CB-SOS = Cognitive Bias Scale of Scales. Sen. = sensitivity. Spec. = specificity. Bold values are those which represent the cutoff score recommended by the authors. If no bolding was present (e.g. CB-SOS-2 and CBSOS-3 in Ingram et al., in press), this missingness indicates that identified values are below reported score ranges. Not all studies included all PAI cognitive overreporting scales and all scales assessed within each study are listed. Armistead-Jehle et al. (2020) provided a number of classification tables, and those listed here are based on the "pass all or fail any" groups.

Method

Participants

Participants were 404 Operations Enduring Freedom, Iraqi Freedom, New Dawn (OEF/OIF/OND)-era Veteran service members assessed in a multi-site, prospective research study conducted by the Neurocognition Laboratory of the VA Mid-Atlantic Mental Illness Research, Education, and Clinical Center. Participants completed the study between 2006 and 2016. The purpose of this study was to evaluate neuropsychological differences among Iraq/Afghanistan-era Veterans with PTSD and/or a history of TBI compared to healthy controls (Shura et al., 2021b). For the current manuscript, all participants were included regardless of diagnostic status, including controls. Initial exclusion criteria were completion of neuropsychological evaluation within six months of study enrollment, active psychotic symptoms, and presence of an active substance use disorder. Additional exclusion criteria for the current analyses were PAI profiles invalidated by non-content-based responding (Inconsistency [INC] > 72 [$n = 15$] and Infrequency [INF] > 74 [$n = 13$]) and a lack of a completed Word Memory Test (WMT, $n = 4$), typically due to computer malfunction. Application of these exclusion criteria resulted in a final analytic sample of $N = 371$.

Institutional Review Board approval was obtained at each of the three Veterans Affairs Medical Center (VAMC) sites located in the Mid-Atlantic region. Participation was voluntary and written informed consent was obtained prior to conduct of any study activities. Participants were identified and recruited from a prior registry study, the Post-Deployment Mental Health study (Brancu et al., 2017) completed on a prior date and the patient population at two sites ([site 1] $n = 46$; [site 2] $n = 264$), and from VAMC Polytrauma/TBI System of Care inpatient, residential, and outpatient rehabilitation programs at a third site ([site 3] $n = 61$). Although all participants were volunteers for the Neurocognition study, all were enrolled in the VA for health care, and a small number dually consented to use their data for clinical purposes at one site (site 2). Individuals were reimbursed for their time and travel. The welfare of human subjects was protected.

Measures

The larger study from which these data derive included 21 cognitive measures, the WMT, and five self-report symptom measures, including the PAI. Specific measures used in the present analysis are described below. All measures were administered in a fixed order and according to test manuals.

Miller forensic assessment of symptoms test (M-FAST; Miller, 2001)

The M-FAST is a brief, structured interview approach to identifying feigning. The measure includes 25 items across seven strategies for identifying response bias, including discrepancies between reported and observed symptoms, extreme symptoms, rare symptom combinations, unusual psychotic symptoms, unusual symptom course, overly negative self-image, and suggestibility. The cutoff score recommended in the manual was used to identify invalid responders in the current

sample (8.7%, $n = 29$). The test manual (Miller, 2001) reports an initial reliability of α = .92.

Word memory test (WMT; Green, 2005)

The WMT presents a series of 20 paired words for two learning trials followed by a forced-choice Immediate Recognition (IR) trial and subsequent Delayed Recognition (DR) trial after 30 minutes. A Consistency Score (CNS) is calculated from the number of items that were correct on both IR and DR trials compared with correct on one trial and incorrect on the other trial. These three scores are the primary "effort" scores for the measure and were used in analyses. When designing the study battery, not all WMT subtests were included given the objectives, time demands, and logistic constraints of the study. Thus, the Genuine Memory Impairment Profile ("dementia profile") cannot be examined. The WMT was scored using the established cutoffs in the manual to identify valid and invalid scores (Green, 2005). For this study, if IR, DR, or CNS was in the invalid score range, the participant was identified as invalid (21.3%; $n = 79$).

Personality assessment inventory (PAI; Morey, 2007)

The PAI is a 344-item, multi-scale, self-report measure of psychiatric symptoms and distress with four main symptom validity scales and several additional scales validated primarily for the detection of psychiatric symptom over-reporting (McCredie & Morey, 2018). Along with validity scales, the PAI contains numerous substantive scales assessing clinical and treatment relevant considerations. In general, the PAI has been robustly studied and validated in Veterans (e.g. Bellet et al., 2018; Blalock et al., 2020; Ingram et al., 2022; Miskey et al., 2015; Shura et al., 2021b; Wallin et al., 2009). The manual reports initial median reliabilities across scales from $\alpha = 0.81$ to $\alpha = 0.86$. The Cognitive Bias Scale and three Scale of Scales were independently created using existing PAI items and scales (Boress et al., 2022; Gaasedelen et al., 2019). Initial reliability for CBS was α = .77; reliabilities were not reported for the three SOS.

Analytic plan

Tests were administered by neuropsychologists and post-doctoral fellows or psychology technicians under supervision of neuropsychologists. The WMT and PAI were administered on identical laptop computers at each of the three VAMC sites. CBS and CB-SOS scales were calculated from the PAI. Participants were then categorized as valid/invalid based on the WMT (PVT) and the M-FAST (SVT). Invalid scores on a single PVT was used as criterion. Of note, many sources recommend invalid scores on two or more measures as criterion. However, this follows the assumption that true cognitive impairment drives the failure. This is exceptionally unlikely with the WMT in our sample that consists of Veterans who have no compelling neurological reason to fail the IR, DR, or CNS subtests (especially given that the majority of moderately to severely neurologically impaired individuals studied in research pass the WMT [see for example, Carone, 2014; Donders & Boonstra, 2007; Green et al., 2009]). Also, the NAN position paper by Bush et al. (2005, p. 422) states: "Performance below established cut-off

scores on *one or more* [emphasis added] well-validated tests designed to measure exaggeration or fabrication of cognitive deficits suggests insufficient effort to do well." Finally, the updated 2021 AACN consensus paper (Sweet et al., 2021, pp. 36–37) notes, "However, to date, there is a lack of consensus as to the best approach to classify the single-fail cases, which in the view of the consensus panel requires additional research." Consequently, we feel like the use of the WMT in isolation is adequate for the current study. This also provides a direct comparison to one set of analyses performed in the Armistead-Jehle et al. (2020) CBS cross-validation (Table 2).

Correlations were computed between all study measures. Mean differences on cognitive bias scales were calculated between M-FAST and WMT valid/invalid groups using independent sample *t*-tests and Cohen's *d* effect size magnitudes. Effects were considered small (0.20–0.49), medium (0.50–0.79), or large (≥0.80), consistent with recommendations of interpretation for clinical sample studies in response style (Rogers, 2008). Medium and large effects indicate clinically meaningful differences (Rosnow et al., 2000) and were determined *a priori* as required for supporting differences between groups. Following mean comparisons, classification accuracy of the CBS and CB-SOS scales was evaluated using area under the curve (AUC) along with sensitivity, specificity, and hit rate across various scale point-estimates. Effective classification was defined by a specificity of at least .90 (Sweet et al., 2021), indicating a low rate of false positives. Large effects for classification accuracy were determined by AUC values greater than .714, which approximate large effect sizes (Cohen, 1988; Rice & Harris, 2005). Data were analyzed using SAS Enterprise Guide and Stata MP version 17.0.

Results

Participant demographics are reported in Table 2. Demographic makeup of the sample was generally male (88.1%) and White (66.1%), with a mean age of 35.0 years and at least some college education (69.6%). Participants were heterogeneous across demographic factors and diagnoses. Participants identified as invalid using the WMT or M-FAST were similar in age to those who did not (*t* tests for age $p > .05$). In contrast, males were more likely to invalidate the WMT compared to females: invalid males = 23.85%; invalid females = 2.27%; χ^2 (1, $n = 371$) = 10.78, Fisher's < .001. Small cell size limited the ability to compare education level across validity groupings. Diagnostic and disability variables are presented in Table 3 by validity status. As would be expected, the M-FAST was related to major depressive disorder (MDD) and posttraumatic stress disorder (PTSD), but not TBI. In contrast, the WMT was related to TBI status and PTSD, but not MDD. Neither current service-connected disability, nor actively filing for disability claim statuses were significantly higher in the invalid groups for both WMT and M-FAST. Significant but weak correlations were observed among percent service-connected disability the four CBS: CBS $r = 0.35$, SOS1 $r = 0.30$, SOS2 $r = 0.39$, SOS3 $r = 0.33$. Additionally, there were no significant differences in WMT invalidity rates across the three sites: site 1, $n = 9$ (19.57%), site 2, $n = 53$ (20.08%), site 3, $n = 17$ (27.87%), χ^2 (2, $n = 371$) = 1.89, $p = .389$, $V = 0.07$. In contrast, there was a significant difference in proportion of invalid M-FAST scores across the three sites: site 1, $n = 12$ (26.67%), site 2, $n = 16$ (7.02%), site 3, $n = 1$ (1.69%), χ^2 (2, $n = 332$) = 22.66, $p < .001$, $V = 0.26$.

Table 2. Sample characteristics (N = 371).

Variable	*M* or *n*	*SD* or %	Min - max
Demographics			
Age	35.19	9.72	19–64
Education			
Incomplete high school	1	0.27%	
GED	6	1.62%	
High school	106	28.57%	
Some college	111	29.92%	
Associate degree	50	13.48%	
College degree	76	20.49%	
Graduate degree	21	5.66%	
Male	327	88.14%	
Race			
White	249	67.12%	
African American	105	28.30%	
Asian	3	0.81%	
Other	14	5.17%	
Hispanic	16	4.31%	
Branch (*n* = 277)			
Army	179	64.62%	
Navy	25	9.03%	
Marines	42	15.16%	
Air Force	29	10.47%	
Coast Guard	2	0.72%	
Service connected disability, yes for any cause (*n* = 328)	188	57.32%	
Filed for disability (*n* = 331)	245	74.02%	
Service connected percentage (*n* = 184)	47.72	28.04	0–100
Deployments			
0	43	11.59%	
1	160	43.13%	
2	89	23.99%	
3 +	79	21.29%	
Lifetime Diagnostic Variables (*N* = 325)			
Bipolar I	4	1.23%	
Bipolar II	3	0.92%	
Major Depressive Disorder	123	37.85%	
Dysthymic Disorder	9	2.77%	
Schizophrenia	0	0.00%	
Panic Disorder	15	4.62%	
Obsessive Compulsive Disorder	4	1.23%	
Posttraumatic Stress Disorder	136	41.85%	
Generalized Anxiety Disorder	14	4.31%	
Adjustment Disorder	7	2.15%	
Traumatic Brain Injury (*n* = 314)			
None	160	50.96%	
Single Mild	63	20.06%	
Multiple Mild	35	11.15%	
Moderate/Severe	56	17.83%	
Outcome Variables			
Personality Assessment Inventory			
Cognitive Bias Scale (CBS)	10.64	5.48	2–27
Scale of Scales 1 (SOS1)	56.30	11.28	39–84
Scale of Scales 2 (SOS2)	0.66	0.15	0.44–1.21
Scale of Scales 2 (SOS2b[sum])	3.99	0.91	2.64–7.25
Scale of Scales 3 (SOS3)	57.35	10.71	39.29–90.86
Word Memory Test (WMT; *n* = 371)			
Immediate Recognition (IR)	92.74	10.99	22.5–100.0
Delayed Recognition (DR)	92.82	11.96	25.0–100.0
Consistency (CNS)	90.55	11.82	42.5–100.0
Invalid	79	21.29%	
Miller Forensic Assessment of Symptoms Test (M-FAST; *n* = 332)			
Total	1.98	3.10	0–21
Invalid	29	7.82%	

Note. Invalidity was determined using the following cutoff scores for the WMT and M-FAST: WMT invalid < 83; M-FAST invalid > 5. For variables with missing data, reduced sample size is noted next to the variable name in column 1. Psychodiagnostic data are per SCID *DSM-IV* diagnoses for those who completed the prior Post-Deployment Mental Health Study.

Table 3. Diagnostic and disability characteristics by validity groupings.

	WMT					M-FAST				
Variable	Valid	Invalid	χ^2	*p*	ES	Valid	Invalid	χ^2	*p*	ES
Lifetime MDD	93/262 (35.50%)	30/63 (47.62%)	3.17	.075	0.10	89/262 (33.97%)	22/26 (84.62%)	25.61	<.001	0.30
Lifetime PTSD	95/262 (36.26%)	41/63 (65.08%)	17.33	<.001	0.23	102/262 (38.92%)	21/26 (80.77%)	16.92	<.001	0.24
TBI			27.92	<.001	0.30			4.28	.232	0.12
None	143/248 (57.66%)	17/66 (25.76%)				130/254 (51.18%)	8/24 (33.33%)			
Single Mild	37/248 (14.92%)	26/66 (39.39%)				52/254 (20.47%)	5/24 (20.83%)			
Multiple Mild	28/248 (11.29%)	7/66 (10.61%)				27/254 (10.63%)	3/24 (12.50%)			
Moderate/Severe	40/248 (16.13%)	16/66 (24.24%)				45/254 (17.72%)	8/24 (33.33%)			
Receiving VA disability	150/263 (57.03%)	38/65 (58.46%)	0.04	.835	0.01	149/261 (57.09%)	18/28 (64.29%)	0.54	.464	0.04
VA disability filed	190/264 (71.97%)	55/67 (82.09%)	2.85	.092	0.09	193/264 (73.11%)	24/29 (82.76%)	1.27	.260	0.07

Note. MDD = major depressive disorder; PTSD = posttraumatic stress disorder; TBI = traumatic brain injury; WMT = Word Memory Test; M-FAST = Miller Forensic Assessment of Symptoms Test; ES = effects size (ϕ or V). Variables were only available for those who completed the prior PDMH study. Diagnoses based on Structured Clinical Interview of *DSM–IV* Axis I Disorders (First et al., 1997). TBI history was established using a self-report measure modified from a questionnaire previously published (Ivins et al., 2003).

Table 4. T tests for CBS and CB-SOS across valid and invalid groups (N = 371).

	WMT				M-FAST			
Scale	Valid *n* = 292	Invalid *n* = 79	*t*	*d*	Valid *n* = 303	Invalid *n* = 29	*t*	*d*
CBS	9.98 (5.13)	13.06 (6.05)	4.55	.55	10.13 (5.22)	16.62 (4.97)	6.45	1.27
SOS1	55.05 (11.08)	60.95 (10.87)	4.22	.54	55.24 (10.77)	68.87 (9.56)	6.60	1.34
SOS2	3.87 (0.86)	4.42 (0.95)	4.93	.61	3.90 (0.86)	5.06 (0.73)	7.08	1.45
SOS3	56.10 (10.43)	61.96 (10.52)	4.42	.56	56.34 (10.30)	69.19 (8.12)	6.54	1.39

Note. WMT = Word Memory Test; M-FAST = Miller Forensic Assessment of Symptoms Test; CBS = Cognitive Bias Scale; SOS = Scale of Scales.
None required Satterthwaite *t* test. All *p* values < .001.

Of the 332 participants who completed both the M-FAST and WMT, most (*n* = 246) passed both. Invalid participants (*n* = 86) were slightly more likely to invalidate the WMT (*n* = 57) than the M-FAST (*n* = 16) and only a minority (17.8%) of invalid participants invalidated both measures (*n* = 13). The relationship between WMT and M-FAST invalidity was statistically significant with a small effect size, which suggests a limited association exists between memory and psychopathology criterion outcomes, χ^2 (*1, n = 332*) = 10.77, *p* = .001, V = .18. For the full sample, mean scores of PAI subscales from which the CBS and CB-SOS were calculated were generally elevated between half and a full standard deviation beyond their normative levels, with SUI serving as the sole exception of elevated mean scores (*M* = 50.25, *SD* = 10.44, range = 10–95). However, observed score ranges for all scales were wide and indicated no issues with restriction of range, which might affect CBS/CB-SOS utility.

Independent samples *t*-tests demonstrated significant, moderate effect differences on CBS, CB-SOS-1, CB-SOS-2, and CB-SOS-3 between WMT-classified groups (Table 4). Independent sample *t*-tests demonstrated large effect size differences for each of the cognitive bias measures between M-FAST-classified groups. Correlations between the CBS and CB-SOS scales and M-FAST and WMT scores were consistent with large relationships for the M-FAST (range *r* = 0.52–0.55) and moderate associations for the WMT (range *r* = −0.25 to −0.35). Pearson correlations between M-FAST Total score, CBS, and each CB-SOS were similar and fell within the large range (*r* = 0.52–0.55). In contrast, correlations between the WMT and CBS/CB-SOS scales were weaker, with moderate values for IR, DR, and CNS scores (*r* = −0.25 to −0.35).

Similarly, ROC analyses (Table 5) produced moderate AUC effects across the WMT-classified groups (AUC = .650 −.676) and large effects across M-FAST-classified groups (AUC = .816 −.854). Scores presented include the maximum sensitivity at specificity of ≥ .90, plus the next higher and lower score. When specificity was set to .90 across the CBS and CB-SOS measures, sensitivity was consistently higher for the M-FAST outcomes than for the WMT: even though M-FAST results were higher, results remained relatively modest (.27–.41).

Discussion

This study evaluated the CBS and CB-SOS scales in a post-deployment Veteran sample, which are recently developed measures of cognitive symptom over-reporting and the only currently available scales for the PAI. We evaluated classification performance

Table 5. Diagnostic accuracy of PAI scales to WMT and M-FAST.

Criterion	Scale	AUC	≥ Cutoff	Sensitivity	Specificity	% Correctly Classified	LR+	LR−
WMT	CBS	.650	17	32.91	86.64	75.20	2.46	0.77
			18	**25.32**	**90.07**	**76.28**	**2.55**	**0.83**
			19	22.78	93.84	78.71	3.70	0.82
	SOS1	.660	73	12.66	89.38	73.05	1.19	0.98
			74	**12.66**	**91.10**	**74.39**	**1.42**	**0.96**
			75	11.39	92.12	74.93	1.45	0.96
	SOS2	.676	5.2	24.05	89.04	75.20	2.19	0.85
			5.3	**22.78**	**90.75**	**76.28**	**2.46**	**0.85**
			5.4	20.25	91.78	76.55	2.46	0.87
	SOS3	.660	71	21.52	88.01	73.85	1.80	0.89
			72	**17.72**	**90.41**	**74.93**	**1.85**	**0.91**
			73	17.72	91.78	76.01	2.16	0.90
M-FAST	CBS	.816	18	41.38	89.18	85.44	3.82	0.66
			19	**41.38**	**92.98**	**88.95**	**5.90**	**0.63**
			20	37.93	93.86	89.49	6.18	0.66
	SOS1	.823	72	41.38	89.47	85.71	3.93	0.66
			73	**27.59**	**90.35**	**85.44**	**2.86**	**0.80**
			74	24.14	91.52	86.25	2.85	0.83
	SOS2	.854	5.2	48.28	89.18	85.98	4.46	0.58
			5.3	**41.38**	**90.35**	**86.52**	**4.29**	**0.65**
			5.4	37.93	91.52	87.33	4.47	0.68
	SOS3	.832	71	41.38	88.30	84.64	3.54	0.66
			72	**37.93**	**90.94**	**86.79**	**4.18**	**0.68**
			73	37.93	92.11	87.87	4.80	0.67

Note. WMT = Word Memory Test; M-FAST = Miller Forensic Assessment of Symptoms Test; CBS = Cognitive Bias Scale; SOS = Scale of Scales.

Bolded values are those which exceed specificity of .90 in this study. SOS1 and SOS3 were rounded to whole numbers; SOS2 and SOS2b rounded to 1 decimal place.

across a variety of potential cutoff scores for these scales using two stand-alone criteria evaluating different aspects of over-reporting (Sherman et al., 2020), which has not been done previously. Our findings suggest that the CBS and, to a lesser extent, CB-SOSs perform well in Veterans when compared to another SVT, but more modestly when compared to a PVT. More specifically, the CBS and CB-SOS perform differently from prior, non-Veteran validation studies and are most like active-duty populations relative to other neuropsychological samples. In addition, they are better at identifying invalid symptom presentation than invalid performance on cognitive testing.

Our findings support use of CBS and CB-SOS-2 within Veteran neuropsychological evaluations to broadly screen invalid presentations. Given the high rate of neuropsychological-related concerns within Veteran populations, and the high base rate of invalid responding on self-report personality measures (Ingram et al., 2020), having scales validated on this unique population with a leading personality inventory is critical for effective neuropsychological practice. Cutoff scores needed to achieve acceptably low false positive rates (specificity ≥ .90) were generally comparable across the PVT and SVT outcome groups, suggesting that regardless of the type of validity (Sherman et al., 2020), the CBS/CB-SOS scales can identify individuals with good false-positive rates (specificity), albeit it with somewhat more variation on false-negative rates (sensitivity).

Cutoff scores for CBS were congruent with prior research (e.g. Armistead-Jehle et al., 2020; Gaasedelen et al., 2019), suggesting that an analogous cutoff score is

needed to achieve minimal standards relative to other populations. In non-Veteran disability litigants (Tylicki et al., 2021), higher cutoffs were necessary for similar performance, which was not replicated within this Veteran sample. In other words, any compensation and pension evaluations were independent of this research evaluation and may not have been ongoing concurrently, possibly producing the observed discrepancy from the Tylicki's medicolegal sample. Similarly, Armistead-Jehle et al. (2020) excluded medical board and disability evaluations from their active-duty sample. Thus, work validating the CBS in Veterans undergoing compensation and pension evaluations and active-duty personnel actively involved in in the medical board process is warranted. The CB-SOS scales (Boress et al., 2022) have undergone notably less cross-validation than their CBS counterpart and, consequently, cutoff score recommendations have a less robust literature for comparison. We found that, relative to their validation, each CB-SOS had a lower cutoff score needed to perform equitably, contrary to our hypothesis. One possible reason for this effect is differences in sample characteristics. Prior studies evaluated the scales in mixed clinical, active duty, and forensic samples, compared to our primary research sample of Veterans.

Although there is a paucity of research on PAI overreporting scales and their association with PVTs (Fokas & Brovko, 2020), classification rates for PAI SVTs have often produced medium effects (Gaasedelen et al., 2019; Morris et al., 2022). We observed that CBS/CB-SOS performance was better on a stand-alone SVT (i.e. M-FAST), relative to a stand-alone PVT (i.e. WMT), with PVT group classification rates and effect sizes comparable to prior work. These findings may naturally reflect the underlying effectiveness of the criterion used to make these classifications. Cognitively focused, mixed-method SVTs in broadband instruments (i.e. those integrating PVT into SVT development; Gaasedelen et al., 2019; Gervais et al., 2007) rely on items being selected in a *post hoc* fashion from an existing pool within an already established instrument, rather than developed in advance as part of an intentional item-pool generation process (Burchett & Bagby, 2022; Rogers & Bender, 2018). It is possible that the item-pool used to create mixed-method embedded SVTs are not broad enough to capture the scope of content relative to cognitive overreporting or to fully take advantage of the relied upon floor effect strategy (i.e. individuals endorsing symptoms/experiences that indicate impairment that is atypical of impaired persons). Thus, across various samples and different broadband measures, overreporting indicators intending to assess distinct constructs perform equitably in their effect sizes, even when distinctive sets of symptoms are targeted in feigning (Morris et al., 2021; Reeves et al., 2022).

Given these results and prior research, symptom feigning may be a broader and more non-specific construct than previously thought; however, cognitive overreporting measures do not contain item content that is exclusive to cognitive symptoms. This criticism is not unique to the PAI and has similarly appeared on scales on other broadband measures (Butcher et al., 2008). The Minnesota Multiphasic Personality Inventory (MMPI) family of instruments Response Bias scale (RBS; Gervais et al., 2007), from which the CBS was patterned (Gaasedelen et al., 2019), has highlighted the inherent overlap in psychopathological response and somatic response as part of suggested interpretations (Gervais et al., 2008, p. 1074). This can be taken to reflect

a broad concern about somatic/cognitively focused over-reporting scales. For example, the CBS includes items from scales assessing depression, anxiety, and treatment rejection (Gaasedelen et al., 2019) and each CB-SOS scale includes numerous clinical/substantive scales that are not exclusive to cognitive symptoms (e.g. depression; Boress et al., 2022).

Revisions to testing measures that aim to expand cognitive overreporting assessment, and to focus on this domain of symptom response (Sherman et al., 2020), may benefit from increased emphasis on the development of cognitively focused items based on *a priori*, empirically based content. Explicit use of validity detection patterns (Rogers & Shuman, 2005) at early developmental phases (e.g. creating specific items that highlight symptom incongruence or symptom rarity; Rogers & Bender, 2018), rather than post-hoc identification of items that may not measure those constructs explicitly is warranted. Well-specified item-pool revision efforts specific to validity testing needs and standards (Sherman et al., 2020; Martin et al., 2015) may not only improve general and longstanding classification difficulties (i.e. low sensitivity), but the distinctiveness of symptom clusters. Even if the overlapping PVT/SVT performance does not resolve entirely as a function of shifted developmental priorities, placing an increased emphasis on validity content development at the test revision stage remains necessary. Broadband measures are widely and historically preferred because of their SVTs (Ben-Porath & Waller, 1992; Russo, 2018), as well as the broader growth in focus on SVT-related research (Sweet et al., 2021). Research on validity scales tends to use *either* PVT or SVT criterion as an outcome, but rarely within the same study. Given the potential for effective performance on the related (but not overlapping) constructs of PVTs and SVTs, the inclusion of distinct criterion measures that assess divergent over-reporting symptom sets (somatic, cognitive, or psychological; Sweet et al., 2021) is also merited.

In addition to issues with the creation of cognitive SVTs noted, there is a tendency to validate cognitive SVTs against PVTs. SVTs and PVTs are different constructs, with subjects invalidating both at rates of only 5 to 11% (Copeland et al., 2016; Shura et al., 2021b). Therefore, using PVTs as criteria for cognitive SVTs might not make sense, and other approaches to validating cognitive SVTs are warranted. As one example, Henry et al. (2014) created the Cognitive Complaints Scale on the MMPI-2 by identifying items with cognitive-based content, as opposed to statistical item identification, which has led to non-cognitive-based item inclusion on scales such as the CBS and RBS. Although this scale is an example of content-based SVT creation, Henry et al. (2014) still evaluated the scale in comparison to PVT status. Given the moderate sensitivity of the CCS to PVT status, the authors concluded that there is a dissociation between PVTs and SVTs. As seen from our results, although the CBS and SOSs were designed to predict PVTs, they better predict feigned psychological symptoms when both PVTs and SVTs are examined. Future research might address this problem by evaluating cognitive SVTs using simulation designs and other SVTs for validation, not PVTs. Regardless, these results further support the independent nature of PVTs and SVTs and highlights the need to assess both in a given evaluation, regardless of the context.

No study is without limitations. We retrospectively analyzed a sample of Veterans who completed the PAI and concurrent PVT/SVT outcomes. We relied on a single

PVT to identify invalid performance, and therefore some individuals may have yielded false negatives given that invalid performance can wax and wane over time, as well as across domains and may be missed if using a single PVT. However, 21.2% of our research-context sample produced invalid scores on the WMT, consistent with the expected rate of failure on a PVT in a Veteran/active-duty sample (Denning & Shura, 2019). Further, prior research with Veterans has found that those with invalid scores on a single PVT show test profiles more similar to those who have invalid scores on two or more PVTs than to those who are valid (Proto et al., 2014); nonetheless, discrepancies with prior CBS studies could be due to our reliance on a single, standalone (though highly sensitive) PVT. Not all individuals completed both the PVT and SVT, so we were unable to produce precise estimates of concurrent failure. Lastly, future research should investigate how neurocognitive disorder diagnosis is associated with the CBS/CB-SOS scales, as well as the grouping criterion measures used within this study since those diagnoses could potentially result in false positive rates, which may explain differential failure rate for criterion measures across TBI severity.

Although generally representative of Veteran populations, our study also over-represents men and White individuals. This limits the generalizability of our results and precluded evaluation of meaningful sex or racial differences. Similarly, this sample was primarily a volunteer research sample; however, a portion of participants at one site and the majority of participants at another site were clinical, which may somewhat mitigate the decreased generalizability related to context. Additionally, the sample showed mixed diagnostic status, and including healthy controls, which may affect generalizability to the greater post-deployment population. The primary research nature of the sample may also explain the lack of relationship between disability status and PVT/SVT failure, but also likely reflects the ubiquitous nature of VA disability among Veterans. Furthermore, diagnostic status was not evaluated on the day of cognitive testing, thus lifetime PTSD and MDD were used to describe the sample for those who completed the PDMH study; given diagnostic status was unknown on the day of testing, such may or may not have affected results. Finally, the use of the M-FAST may limit the strength of conclusions, given concerns regarding the psychometrics of the measure and that the test was designed to be a screener (Wolf et al., 2020); however, a meta-analysis on the M-FAST found a very large effect size when comparing validity groups, and with adequate psychometrics in independent groups (Detullio et al., 2019). These limitations notwithstanding, our study provides an expanded body of support for the CB and CB-SOS scales with guidance for their use within Veteran neuropsychological evaluations.

Acknowledgments

This work was supported by the Salisbury VA Health Care System, VA Mid-Atlantic (VISN 6) Mental Illness Research, Education, and Clinical Center (MIRECC), and the Dwight D. Eisenhower Veteran Affairs Medical Center. The authors declare no conflicts of interest, financial or otherwise. We would like to thank Dr. Larry Tupler for his work on establishing and previously directing the VA Mid-Atlantic MIRECC Neurocognition lab.

Disclaimer

The views, opinions and/or findings contained in this article are those of the authors and should not be construed as an official Veterans Affairs, Defense Health Agency, or Department of Defense position, policy or decision, unless so designated by other official documentation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Robert D. Shura http://orcid.org/0000-0002-9505-0080
Paul B. Ingram http://orcid.org/0000-0002-5409-4896
Holly M. Miskey http://orcid.org/0000-0002-5139-4586
Sarah L. Martindale http://orcid.org/0000-0002-0059-3454
Jared A. Rowland http://orcid.org/0000-0002-3655-3411
Patrick Armistead-Jehle http://orcid.org/0000-0002-6784-1432

References

Armistead-Jehle, P., Gervais, R. O., & Green, P. (2012). Memory complaints inventory results as a function of symptom validity test performance. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *27*(1), 101–113.

Armistead-Jehle, P., Ingram, P. B., & Morris, N. M. (2020). Personality assessment inventory cognitive bias scale: Validation in a military sample. *Archives of Clinical Neuropsychology*, *35*(7), 1154–1161. https://doi.org/10.1093/arclin/acaa049

Bellet, B. W., McDevitt-Murphy, M. E., Thomas, D. H., & Luciano, M. T. (2018). The utility of the Personality Assessment Inventory in the assessment of posttraumatic stress disorder in OEF/OIF/OND veterans. *Assessment*, *25*(8), 1074–1083. https://doi.org/10.1177/1073191116681627

Ben-Porath, Y. S., & Waller, N. G. (1992). "Normal" personality inventories in clinical assessment: General requirements and the potential for using the NEO Personality Inventory. *Psychological Assessment*, *4*(1), 14–19. https://doi.org/10.1037/1040-3590.4.1.14

Blalock, D. V., Wilson, S. M., Dedert, E. A., Clancy, C. P., Hertzberg, M. A., Beckham, J. C., & Calhoun, P. S. (2020). The Personality Assessment Inventory Alcohol Scale in veterans with PTSD: Convergent and discriminant relations with the alcohol use disorders identification test. *Assessment*, *27*(4), 719–727. https://doi.org/10.1177/1073191118824661

Boress, K., Gaasedelen, O. J., Croghan, A., Johnson, M. K., Caraher, K., Basso, M. R., & Whiteside, D. M. (2022). Replication and cross-validation of the personality assessment inventory (PAI) cognitive bias scale (CBS) in a mixed clinical sample. *The Clinical Neuropsychologist*, *36*(7), 1860–1877. https://doi.org/10.1080/13854046.2021.1889681.

Brancu, M., Wagner, H. R., Morey, R. A., Beckham, J. C., Calhoun, P. S., Tupler, L. A., Marx, C. E., Taber, K. H., Hurley, R. A., Rowland, J., McDonald, S. D., Hoerle, J. M., Moore, S. D., Kudler, H. S., Weiner, R. D., & Fairbank, J. A. (2017). The Post-Deployment Mental Health (PDMH) study and repository: A multi-site study of US Afghanistan and Iraq era veterans. *International Journal of Methods in Psychiatric Research*, *26*(3), e1570. https://doi.org/10.1002/mpr.1570

Burchett, D., & Bagby, R. M. (2022). Assessing negative response bias: A review of the noncredible overreporting scales of the MMPI-2-RF and MMPI-3. *Psychological Injury and Law*, *15*(1), 22–36. https://doi.org/10.1007/s12207-021-09435-9

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*(4), 419–426. https://doi.org/10.1016/j.acn.2005.02.002

Butcher, J. N., Gass, C. S., Cumella, E., Kally, Z., & Williams, C. L. (2008). Potential for bias in MMPI-2 assessments using the Fake Bad Scale (FBS). *Psychological Injury and Law*, *1*(3), 191–209. https://doi.org/10.1007/s12207-007-9002-z

Carone, D. A. (2014). Young Child With Severe Brain Volume Loss Easily Passes the Word Memory Test and Medical Symptom Validity Test: Implications for Mild T BI. *The Clinical Neuropsychologist*, *28*(1), 146–162. https://doi.org/10.1080/13854046.2013.861019

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, (2nd ed.). Lawrence Erlbaum.

Copeland, C. T., Mahoney, J. J., Block, C. K., Linck, J. F., Pastorek, N. J., Miller, B. I., Romesser, J. M., & Sim, A. H. (2016). Relative utility of performance and symptom validity tests. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *31*(1), 18–22. https://doi.org/10.1093/arclin/acv065

Denning, J. H., & Shura, R. D. (2019). Cost of malingering mild traumatic brain injury-related cognitive deficits during compensation and pension evaluations in the Veterans Benefits Administration. *Applied Neuropsychology. Adult*, *26*(1), 1–16. https://doi.org/10.1080/23279095.2017.1350684

Detullio, D., Messer, S. C., Kennedy, T. D., & Millen, D. H. (2019). A meta-analysis of the Miller Forensic Assessment of Symptoms Test (M-FAST). *Psychological Assessment*, *31*(11), 1319–1328. https://doi.org/10.1037/pas0000753

DeViva, J. C., & Bloem, W. D. (2003). Symptom exaggeration and compensation seeking among combat veterans with posttraumatic stress disorder. *Journal of Traumatic Stress*, *16*(5), 503–507. https://doi.org/10.1023/A:1025766713188

Donders, J., & Boonstra, T. (2007). Correlates of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury*, *21*(3), 319–326. 10.1080/02699050701253129

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured clinical interview for DSM–IV–TR axis I disorders, research version, patient ed. (SCID-I/P)*. Biometrics Research, New York State Psychiatric Institute.

Fokas, K. F., & Brovko, J. M. (2020). Assessing symptom validity in psychological injury evaluations using the MMP I-2-RF and the PAI: An updated review. *PsychologicalInjury and Law*, *13*(4), 370–382. https://doi.org/10.1007/s12207-020-09393-8

Freeman, T., Powell, M., & Kimbrell, T. (2008). Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. *Psychiatry Research*, *158*(3), 374–380. https://doi.org/10.1016/j.psychres.2007.04.002

Gaasedelen, O. J., Whiteside, D. M., Altmaier, E., Welch, C., & Basso, M. R. (2019). The construction and initial validation of the cognitive bias scale for the personality assessment inventory. *The Clinical Neuropsychologist*, *33*(8), 1467–1484. https://doi.org/10.1080/13854046.2019.1612947

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment*, *14*(2), 196–208.

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2008). Differential sensitivity of the Response Bias Scale (RBS) and MMPI-2 validity scales to memory complaints. *The Clinical Neuropsychologist*, *22*(6), 1061–1079.

Green, P. (2005). *Green's Word Memory Test for Windows user's manual*. Edmonton, Canada: Green's Publishing Inc.

Green, P., Flaro, L., & Courtney, J. (2009). Examining false positives on the Word Memory Test in adults with mild traumatic brain injury. *Brain Injury*, *23*(9), 741–750. https://doi.org/10.1080/02699050903133962

Hawes, S. W., & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: A meta-analytic review. *Psychological Assessment*, *21*(1), 112–124.

Henry, G. K., Heilbronner, R. L., Mittenberg, W., Hellemann, G., & Myers, A. (2014). Development of the MMPI-2 cognitive complaints scale as an embedded measure of symptom validity. *Brain Injury*, *28*(3), 357–363. https://doi.org/10.3109/02699052.2013.865272

Ingram, P. B., Armistead-Jehle, P., Morris, N. M., & Herring, T. T. (in press). Cross validation of the Personality Assessment Inventory (PAI) Cognitive Bias Scale of Scale (CBSOS) over-reporting indicators in a military sample.

Ingram, P. B., Morris, N. M., Golden, B., Youngren, W. A., Fulton, J. A., & Sharpnack, J. (2022). The influence of service era: Comparing personality assessment inventory (PAI) scale scores within a posttraumatic stress disorder treatment clinic (PCT). *Journal of Clinical Psychology in Medical Settings*, *29*, 624–635. https://doi.org/10.1007/s10880-021-09812-1

Ingram, P. B., Sharpnack, J. D., Mosier, N. J., & Golden, B. L. (2021). Evaluating symptom endorsement typographies of trauma-exposed veterans on the Personality Assessment Inventory (PAI): A latent profile analysis. *Current Psychology*, *40*(11), 5267–5277. https://doi.org/10.1007/s12144-019-00486-5

Ingram, P. B., Tarescavage, A. M., Ben-Porath, Y. S., & Oehlert, M. E. (2020). Patterns of MMPI-2-Restructured Form (MMPI-2-RF) validity scale scores observed across Veteran Affairs settings. *Psychological Services*, *17*(3), 355–362.

Ivins, B. J., Schwab, K. A., Warden, D., Harvey, L. T. C. S., Hoilien, M. A. J. M., Powell, C. O. L. J., Johnson, C. S. M. E. W., & Salazar, A. M. (2003). Traumatic brain injury in U.S. Army paratroopers: Prevalence and character. *The Journal of Trauma*, *55*(4), 617–621.

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of north American professions. *The Clinical Neuropsychologist*, *29*(6), 741–776. https://doi.org/10.1080/13854046.2015.1087597

McCredie, M. N., & Morey, L. C. (2018). Evaluating new supplemental indicators for the personality assessment inventory: Standardization and cross-validation. *Psychological Assessment*, *30*(10), 1292–1299. http://dx.doi.org/10.1037/pas0000574

Miller, H. A. (2001). *Miller Forensic Assessment of Symptoms Test: Professional Manual*. Psychological Assessment Resources, Inc.

Miskey, H. M., Shura, R. D., Yoash-Gantz, R. E., & Rowland, J. A. (2015). Personality assessment inventory profiles of veterans: Differential effects of mild traumatic brain injury and psychopathology. *Brain Imaging and Behavior*, *9*(3), 461–471. https://doi.org/10.1007/s11682-015-9391-7

Morey, L. C. (2007). *Personality assessment inventory*, (2nd ed.). Psychological Assessment Resources, Inc.

Morris, N. M., Ingram, P. B., & Armistead-Jehle, P. (2022). Relationship of personality assessment inventory (PAI) over-reporting scales to performance validity testing in a military neuropsychological sample. *Military Psychology*, *34*(4), 484–493. https://doi.org/10.1080/08995605.2021.2013059

Morris, N. M., Mattera, J., Golden, B., Moses, S., & Ingram, P. B. (2021). Evaluating the performance of the MMPI-3 over-reporting scales: Sophisticated simulators and the effects of comorbid conditions. *The Clinical Neuropsychologist*, 1–9. https://doi.org/10.1080/13854046.2021.1968037

Mozley, S. L., Miller, M. W., Weathers, F. W., Beckham, J. C., & Feldman, M. E. (2005). Personality Assessment Inventory (PAI) profiles of male veterans with combat-related posttraumatic stress disorder. *Journal of Psychopathology and Behavioral Assessment*, *27*(3), 179–189. https://doi.org/10.1007/s10862-005-0634-6

Pearman, A. (2009). Predictors of subjective memory in young adults. *Journal of Adult Development*, *16*(2), 101–107. https://doi.org/10.1007/s10804-009-9063-1

Proto, D. A., Pastorek, N. J., Miller, B. I., Romesser, J. M., Sim, A. H., & Linck, J. F. (2014). The dangers of failing one or more performance validity tests in individuals claiming mild traumatic brain injury-related postconcussive symptoms. *Archives of Clinical Neuropsychology*, *29*(7), 614–624. https://doi.org/10.1093/arclin/acu044

Ray, C. L. (2017). Practical use of MMPI-2-RF validity indicators in VA compensation and pension examination. *Psychological Injury and Law*, *10*(3), 223–233. https://doi.org/10.1007/s12207-017-9289-3

Reeves, C. K., Brown, T. A., & Sellbom, M. (2022). An examination of the MMPI-3 validity scales in detecting overreporting of psychological problems. *Psychological Assessment*, *34*(6), 517–527. https://doi.org/10.1037/pas0001112

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and Human Behavior*, *29*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7

Rogers, R. E. (2008). *Clinical assessment of malingering and deception*. Guilford Press.

Rogers, R., & Bender, S. D. (2018). *Clinical assessment of malingering and deception 4th ed*. (R. Rogers & S. D. Bender (Eds.)). The Guilford Press.

Rogers, R., & Shuman, D. W. (2005). *Fundamentals of forensic practice: Mental health and criminal law*. Springer.

Rosnow, R. L., Rosenthal, R., & Rubin, D. B. (2000). Contrasts and correlations in effect-size estimation. *Psychological Science*, *11*(6), 446–453.

Russo, A. C. (2013). Ethical, legal, and risk management considerations in neuropsychological assessments of veterans. *Psychological Injury and Law*, *6*(1), 21–30. https://doi.org/10.1007/s12207-013-9145-z

Russo, A. C. (2018). A practitioner survey of Department of Veterans Affairs Psychologists who provide neuropsychological assessments. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *33*(8), 1046–1059. https://doi.org/10.1093/arclin/acx139

Sherman, E. M., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *35*(6), 735–764.

Shura, R. D., Denning, J. H., Armistead-Jehle, P., & Axelrod, B. N. (2021a). Validity testing within the VA and active duty settings. In R. W. Schroeder & P. K. Martin. (Eds.), *Neuropsychological validity testing in clinical settings: A handbook for non-forensic practice*. Guilford Press.

Shura, R. D., Yoash-Gantz, R. E., Pickett, T. C., McDonald, S. D., & Tupler, L. A. (2021b). Relations among performance and symptom validity, mild traumatic brain injury, and posttraumatic stress disorder symptom burden in postdeployment veterans. *Psychological Injury and Law*, *14*(4), 257–268. https://doi.org/10.1007/s12207-021-09415-z

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., … & Suhr, J. A. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *35*(6), 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

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*(1), 71–83. https://doi.org/10.1037/pas0000968

Wallin, M. T., Wilken, J., Alfaro, M. H., Rogers, C., Mahan, C., Chapman, J. C., Fratto, T., Sullivan, C., Kang, H., & Kane, R. (2009). Neuropsychologic assessment of a population-based sample of Gulf War veterans. *Cognitive and Behavioral Neurology: Official Journal of the Society for Behavioral and Cognitive Neurology*, *22*(3), 155–166. https://doi.org/10.1097/WNN.0b013e3181b278e8

Wolf, E. J., Ellickson-Larew, S., Guetta, R. E., Escarfulleri, S., Ryabchenko, K., & Miller, M. W. (2020). Psychometric performance of the Miller Forensic Assessment of Symptoms Test (M-FAST) in veteran PTSD assessment. *Psychological Injury and Law*, *13*(3), 284–302. https://doi.org/10.1007/s12207-020-09373-y

Worthen, M. D., & Moering, R. D. (2011). A practical guide to conducting VA compensation and pension exams for PTSD and other mental disorders. *Psychological Injury and Law*, *4*(3–4), 187–216. https://doi.org/10.1007/s12207-011-9115-2