Cross validation of the Personality Assessment Inventory (PAI) Cognitive Bias Scale of Scales (CB-SOS) over-reporting indicators in a military sample

Paul B. Ingram, Patrick Armistead-Jehle, Tristan T. Herring & Nicole M. Morris

Published online: 05 Jan 2023.

Submit your article to this journal

View related articles

View Crossmark data

Check for updates

Cross validation of the Personality Assessment Inventory (PAI) Cognitive Bias Scale of Scales (CB-SOS) over-reporting indicators in a military sample

Paul B. Ingram [a,b], Patrick Armistead-Jehle [c], Tristan T. Herring [a], and Nicole M. Morris [a]

aDepartment of Psychological Sciences, Texas Tech University, Lubbock, Texas, USA; bDwight D. Eisenhower Veteran Affairs Medical Center, Eastern Kansas Veteran Healthcare System, Leavenworth, Kansas, USA; cMunson Army Health Center, Fort Leavenworth, Kansas, USA

ABSTRACT

Following the development of the Cognitive Bias Scale (CBS), three other cognitive over-reporting indicators were created. This study cross-validates these new Cognitive Bias Scale of Scales (CB-SOS) measurements in a military sample and contrasts their performance to the CBS. We analyzed data from 288 active-duty soldiers who underwent neuropsychological evaluation. Groups were established based on performance validity testing (PVT) failure. Medium effects (d = .71 to .74) were observed between those passing and failing PVTs. The CB-SOS scales have high specificity (≥.90) but low sensitivity across the suggested cut scores. While all CB-SOS were able to achieve .90, lower scores were typically needed. CBS demonstrated incremental validity beyond CB-SOS-1 and CB-SOS-3; only CB-SOS-2 was incremental beyond CBS. In a military sample, the CB-SOS scales have more limited sensitivity than in its original validation, indicating an area of limited utility despite easier calculation. The CBS performs comparably, if not better, than CB-SOS scales. CB-SOS-2's differences in performance in this study and its initial validation suggest that its psychometric properties may be sample dependent. Given their ease of calculation and relatively high specificity, our study supports the interpretation of elevated CB-SOS scores indicating those who are likely to fail concurrent PVTs.

What is the public significance of this article?—This article provides psychologists working conducting neuropsychological evaluations on military personnel information to ensure effective testing conclusions.

Introduction

Best practice guidelines for psychological assessment require that validity testing be integrated into evaluations so that psychologists can ensure that the obtained data are an accurate reflection of patient effort and symptom experience (e.g., Bush et al., 2005; Heilbronner et al., 2009; Martin et al., 2015; Sweet et al., 2021). Such testing is typically conceptualized to measure constructs of either performance (PVT) or symptom (SVT) validity. PVTs and SVTs both assess response validity; however, they do so using distinct approaches. More specifically, Performance Validity refers to validity of test performance, while symptom validity refers to the validity of symptom report

(Larrabee, 2012). While these constructs hold conceptual overlap, research has shown them to be only moderately related (Armistead-Jehle et al., 2012; Martin et al., 2015; Pearman, 2009; Sweet et al., 2021). Despite the incomplete relationship between these approaches, psychologists must successfully integrate this information. This can be a challenging task as the constructs are not perfectly aligned (Harwood et al., 2011).

While methods of detection in performance validity testing consist of several approaches (e.g., force-choice, performance curve, etc.; for an outline of these approaches in clinical practice, see, Rogers & Bender, 2018), generally they work to identify invalid responding through comparisons to normative performances. For instance, a test respondent given a PVT may be assessed to determine if their answers are less accurate than chance (forced-choice strategy) or item-difficulty may be compared to an examinee's average performance over time (performance curve). Accordingly, PVT results provide determinations about the probability of credible cognitive testing response based on observed cognitive

task performance. In contrast to PVTs, symptom validity testing evaluates perceived psychological, somatic, and cognitive symptoms. SVTs often rely on the use of infrequently endorsed symptom patterns, but also use approaches such as infrequent symptom combinations (Rogers & Bender, 2018).

Broadband measures are psychological assessment instruments that examine a wide spectrum of symptom complaints and include embedded validity indicators. They contain scoring and item information that is protected and restricted in who may access it, making broadband assessment distinct from mental health screeners (e.g., Patient Health Questionnaire-9 [PHQ-9]) which are openly accessible and rather face valid. The most commonly administered broadband personality instruments are the Minnesota Multiphasic Personality Inventory (MMPI) family of instruments and the Personality Assessment Inventory (PAI; see, Ingram et al., 2020; Wright et al., 2016). These broadband measures meet a high standard of court-admissible evidence because of the restricted nature of their items and item scoring, as well as their expansive normative and validity data (e.g., Ben-Porath, 2012).

Given their incorporation of validity scales to ensure the highest quality assessment data, broadband measures of psychopathology are often administered alongside cognitive testing. As such, the successful use of broadband personality measures relies on the ability to effectively integrate data from their clinical and validity scales into other forms of testing evidence (e.g., self-report, other-report, other psychological testing data). These broadband measures are also the primary home of SVTs. In a study of assessment practices within the Department of Veterans Affairs, these broadband measures are the most frequently relied upon symptom validity measures given during neuropsychological evaluations (Russo, 2018; Young et al., 2016). Such practices also correspond to the Veteran Affairs (VA)/Department of Defense (DOD) evidence-based clinical practice guidelines for assessment, underscoring the necessity of advancing evidence-based practice with these measures within Active-Duty populations. Taken together and given the wide use of these broadband measures into evidence-based practice guidelines, such as those for traumatic brain injury (Department of Veteran Affairs, 2022), the effectiveness of the personality assessment SVTs is critical to effective diagnostic practice.

Several different approaches are used when constructing SVTs for broadband personality measures. Endorsement infrequency is a particularly common approach when the goal is to assess invalid psychopathology symptoms (Rogers & Bender, 2018; Sweet et al., 2021); however, assessment of cognitive complaints relies on a different approach. "Mixed method" SVTs (i.e., those integrating PVT into SVT development) are a leading approach to SVT cognitive symptom assessment. Despite the distinctiveness of SVTs and PVTs (Martin et al., 2015), this method has been successful in detecting this form of invalid responding (Burchett & Bagby, 2021; Sharf et al., 2017). In mixed-PVT-SVT approaches, candidate SVT scales are validated through the use of external PVT criterion (e.g., pass and fail groups are based on PVT tests, rather than the infrequency of item's endorsement or other external criteria). This mixed-method approach (SVTs validation based on PVT criterion) has shown substantial promise and can be described conceptually as a "boot-strapped" floor effect strategy (see Burchett & Bagby, 2021). The term boot-strapped floor effect is the best conceptual description of these SVTs because items are selected for a scale only if they effectively distinguish between those who pass and fail concurrent PVT testing and have among the best signal detection for over-reported or misrepresented cognitive symptom experiences (Ingram et al., 2022; Sharf et al., 2017). Even when validity scales specifically designed to detect cognitive symptom misrepresentation are not available, broadband personality measures are still used to assess symptom presentation (Keiski, 2007; Till et al., 2009).

The Personality Assessment Inventory (PAI; Morey, 2007) is one of the most frequently used broadband measures of psychopathology (e.g., Ingram et al., 2022); however, it has historically lacked validity scales assessing cognitive symptom over-reporting. Fortunately, recent research has created PAI scales for this purpose. Addressing the lack of cognitive over-reporting indicators for the PAI, Gaasedelen et al. (2019) developed the 10-item Cognitive Bias Scale (CBS) in a mixed neuropsychological sample using a mixed PVT-SVT approach. The CBS uses items which differentiate those who passed and failed PVTs, analogous to the method used on the MMPI-2/RF/3 Response Bias scale (RBS; see, Gervais et al., 2007). In its validation, the CBS outperformed existing symptom validity measures on the PAI (e.g., MAL, NIM, etc.) in predicting PVT failure. The CBS also demonstrated large effect differences (i.e., $d \geq .8$; Cohen, 1988) between those who passed and failed external performance validity testing with acceptable classification accuracy at a cut-score of 16 (sensitivity = .37, specificity = .90, positive and negative predictive powers [PPV and NPV, respectively] = .70).

To contextualize these findings, sensitivity is the percentage of true positives, and specificity is the percentage of true negatives. Likewise, PPV is associated with the probable likelihood that an individual with a positive test result (i.e., meeting or exceeding a specific score on

CBS) is truly feigning, while NPP is the likelihood that they are not misrepresenting cognitive symptoms. While acceptable classification was observed at a CBS score of 16, A score of 19 was preferred by Gaasedelen and colleagues as it produced improved sensitivity and specificity, closer to those observed in the MMPI-2/RF/3 RBS during its validation. The score of 19 still met the .90 specificity threshold, which is standard for validity scales (Sweet et al., 2021). The CBS was subsequently cross validated in an active-duty military sample and achieved comparable performance at a cut score of 16 where specificity was .92, sensitivity = .55, PPV = .59, and NPV = .83 (Armistead-Jehle et al., 2020). In contrast to the CBS' initial validation, a cut score of 19 produced minimal gains in specificity and notably smaller sensitivity values. Thus, the use of the CBS in active-duty military provides evidence of distinct performance compared to outpatient neuropsychological samples.

Shortly after CBS was developed, Boress et al. (2021) constructed three alternative scales designed to detect over-reported cognitive symptoms, termed the Cognitive Bias Scale of Scales (CB-SOS). As with the CBS, Boress et al. (2021) used a mixed neuropsychological sample for the development of these cognitive validity scales. However, rather than an item-level mixed PVT-SVT approach, the CB-SOS component scales were selected based on prior work showing that those scales differed in biased and unbiased responding patterns on external effort tests (Whiteside et al., 2011). Thus, the CB-SOS scales build on the mixed PVT-SVT approaches used to detect invalid cognitive symptoms by relying on scale-level data. Because they use scales, rather than items, they are also easier for clinicians to calculate.

CB-SOS-1 is the average of six PAI scale T-scores (NIM [Negative Impression Management, an Over-reporting Validity scale], SOM [Somatic Complaints, a substantive clinical scale], DEP [Depression, a substantive clinical scale], ANX [Anxiety, a substantive clinical scale], SCZ [Schizophrenia, a substantive clinical scale], and SUI [Suicidal Ideation, a treatment concerns scale]). Conceptualized within contemporary models of psychopathology, these scales assess not only cognitive concerns but also internalization, thought disorder, somatic experiences, and response validity (Kotov et al., 2017). CB-SOS-2 was derived using the average of these same six scales; however, each scale was multiplied by a beta weight derived from a logistic regression in which PAI scales predicted failure on a PVT. CB-SOS-3 was calculated by averaging PAI T-scores of various validity, clinical, and clinical sub-scales (NIM, SCZ, SOM-C [Somatic Complaints-Conversion], SOM-S [Somatic Complaints-Somatization], DEP-P [Depression-Physiological], ANX-P [Anxiety-

Physiological], and PAR-R [Paranoia-Resentment]). Most, but not all, of the scales for the CB-SOS are associated with physical sensations and experiences (i.e., SOM-C, ANX-P, DEP-P, and SOM-S). In their initial validation (see, Boress et al., 2021), each CB-SOS scale demonstrated large classification effects (Area Under the Curve [AUC] of .72 to .75). At recommended cut scores (CB-SOS-1 = T78, CB-SOS-2 = 5.3, CB-SOS-3 = T74), Boress et al. (2021) noted high specificity (.90, .90, and .92, respectively) and lower, modest sensitivity (.29, .41, and .38, respectively). Pass and fail PVT groups had large magnitude differences (d = .81 [CB-SOS-1] to 1.00 [CB-SOS-3]). While these results are promising, in this study the CB-SOS were not contrasted with CBS. As such, it was not possible to contextualize the relative performance of those scales to the CBS, despite CBS being the only PAI scale designed to assess cognitive symptom over-reporting.

Recently, a study published in a sample of Veterans (Shura et al., 2022) provides the first replication of the CB-SOS scales. In this study, the authors used a retrospective database to determine groups of individuals with failed cognitive validity testing (PVT failure) and a group of individuals with failed external symptom validity testing (SVT failure). These groups were then independently contrasted with respondents on the PAI who did not have any failures on concurrently administered validity tests, either PVT or SVT. They found that the CB-SOS scores required to obtain the necessary classification accuracy were lower in Veterans than in the initial validation of those scales (Boress et al., 2021). Thus, the effectiveness of the CB-SOS scales, much like CBS (see Armistead-Jehle et al., 2020), differs when used with those who have military service. Such differences in effectiveness may be due to the sample-dependent influence of using "boot-strapped" floor effects used to create the CBS and CB-SOS scales or may exist as a broader function of the population being evaluated. Sample-dependent patterns are particularly likely when base rates of cognitive difficulty and injury vary between populations, which are substantially more frequent in military populations (e.g., Ratcliffe et al., 2022).

The primary aim of this study is to cross-validate the CB-SOS scales in a sample of active-duty military personnel. A related aim is to contrast CB-SOS performance with the CBS. Given their similar shared developmental processes (e.g., content selected based on performance in predicting PVT failure), we hypothesized that each CB-SOS would adequately discriminate between valid versus invalid groups and that performances on these scales would be commensurate with CBS. We anticipate medium-to-large effects (d > .5; Cohen, 1988) and that Active-Duty personnel would produce distinct patterns requiring cut-score

interpretations, which are different from either outpatient neuropsychological clinic clients (Boress et al., 2021) and from Veterans seen for neuropsychological evaluation with the VA (Shura et al., 2022).

Methods

Participants

We excluded potential participants who were undergoing or pending medical board evaluation ($n = 7$) or temporary disability evaluation ($n = 4$). Participants were also excluded if the PAI non-content-based invalid responding scales exceeded recommended cut scores (INF ≥ 75 [$n = 6$] or ICN ≥ 72 [$n = 3$]; see, Morey, 2007). Consistent with methods commonly used in over-reporting scale studies, we did not exclude participants for invalidation of other content-based over-reporting measures because it is common for individuals to demonstrate elevated scale scores simultaneously across over-reporting indicators of broadband personality measures (e.g., Hawes & Boccaccini, 2009; Ingram et al., 2020; Morris et al., 2022). Broad exclusion based on standard PAI validity scales could cause an attenuated distribution of those within the failed PVT group and bias interpretation of CB-SOS scale classification because of mixed responding approach tendencies (Sweet et al., 2021).

Following exclusions, this study included 288 Army Active-Duty soldiers referred for evaluation to an outpatient neuropsychology clinic at a U.S. Army Health Center. The average age of the sample was 38.9 years (SD = 7.9), with an average education of 16 years (SD = 2.4). The sample was predominantly male 84%), white (70.5%), and contained more officers (61.1%) than enlisted (38.9%) service members.

Of those with a history of mild traumatic brain injury (mTBI; $n = 161$), the most recent head injury was fairly remote (Number of Months M = 119.5, SD = 98.7). Diagnosis of mTBI was defined via the DoD/VA criteria, which includes loss of consciousness 30 min or less; post-traumatic amnesia 24 h or less; self-reported alteration of consciousness/mental state lasting up to 24 hours; or Glasgow Coma Scale score ≥13. This sample did not contain any participants with more severe head injury history, and none of the sample had evidence of other neurological conditions that would explain testing performance. Concussion diagnoses were based on the clinical interview and a medical records review. Nonmutually exclusive psychiatric diagnoses include post-traumatic stress disorder (9.8%), anxiety disorder (27.4%), and depressive disorder (10.9%) and were based on the totality of data gathered during the

psychological evaluation (e.g., record review, semi-structured diagnostic interview, and psychological testing). Approximately 34% of the sample had no psychiatric diagnosis. Diagnoses were made by the second author (a board-certified neuropsychologist) based on the totality of the neuropsychological assessment.

Procedures and measures

Patients were tested by a trained neuropsychology technician under the supervision of the second author. Participants were administered a battery of neuropsychological tests, which included the Medical Symptom Validity Test (MSVT) and Non-verbal Medical Symptom Validity Test (NV-MSVT), as well as the Personality Assessment Inventory (PAI). Prior to the publication of the CB-SOS, these data was used to cross validate CBS within a military population (Armistead-Jehle et al., 2020).

Medical Symptom Validity Test

The MSVT (Green, 2004) is a brief automated verbal memory screening with several subtests designed to measure performance validity. In addition to data presented in the manual, several studies have demonstrated the utility of this measure in the discrimination between those with genuine memory impairment and those simulating impairment in a range of patient samples (see, Carone, 2009 for review). In this sample, 67 participants (94.4% of PVT failure group) failed the MSVT using cut scores defined in the test manual.

Non-verbal Medical Symptom Validity Test

The NV-MSVT (Green, 2008) is a brief automated non-verbal memory screening with several subtests designed to measure performance validity. The NV-MSVT as a measure of performance validity has been widely validated (see, Wager & Howe, 2010 for review). In this sample, 41 participants (57.7% of PVT failure group) failed the NV-MSVT using cut scores defined in the test manual.

Personality Assessment Inventory

The PAI (Morey, 2007) is a measure of personality and emotional functioning consisting of 344 items answered on a 4-point Likert-type format (F = false, not at all true; ST = slightly true; MT = mainly true; and VT = very true). Among PAI scales are four validity scales (i.e., Inconsistency [ICN], Infrequency [INF], Negative Impression Management [NIM], and Positive Impression Management [PIM]). Several supplemental validity scales have also been developed (Hawes & Boccaccini, 2009), of which the CB-SOS and CBS remain

the only ones explicitly designed to measure invalid cognitive symptom endorsement on the PAI.

The PAI also includes 11 clinical scales, which serve to measure various types of psychopathologies (Somatic Complaints [SOM], Anxiety [ANX], Anxiety-related Disorder [ARD], Depression [DEP], Mania [MAN], Paranoia [PAR], Schizophrenia [SCZ], Borderline Features [BOR], Antisocial Features [ANT], Alcohol Problems [ALC], and Drug Problems [DRG]). Most of these primary clinical scales are comprised of several subscales, providing for more nuanced interpretation of symptom patterns. For instance, DEP (Depression) is comprised of three non-overlapping subscales, including Affective (DEP-A), Cognitive (DEP-C), and Physiological (DEP-P) concerns. Additionally, there are five treatment consideration scales (Aggression [AGG], Suicidal Ideation [SUI], Stress [STR], Nonsupport [NON], and Treatment Rejection [RXR]) and two scales, which measure interpersonal functioning (Dominance [DOM] and Warmth [WRM]). In general, the PAI has demonstrated good internal consistency and criterion validity in military (e.g., Bellet et al., 2018; Van Voorhees et al., 2014) and clinical samples (Slavin-Mulford et al., 2012).

Statistical analyses

Participants were administered the same test battery and grouped into PVT pass (passed both PVTs [$n = 207$]) or PVT fail (failed the NV-MSVT and/or MSVT [$n = 81$]). Independent sample t-tests examined between-group mean score differences across CB-SOS and CBS scores and Cohen's d effect sizes were used to quantify the magnitude of these differences. Effect sizes are defined as small (.5 > d > .2), medium (.8 > d ≥ .5), and large (.8 ≥ d) effects (Cohen, 1988). We also identified clinically meaningful differences for between-group comparisons as those with at least a medium effect (i.e., 5 T-score points; Rosnow et al., 2000). Confidence intervals for effect size metrics were also computed.

Area Under the Curve (AUC) is an overall estimation of classification accuracy for a scale, across all potential scale scores observed and ranges in magnitude from 0 to 1.0, with effects at or exceeding .71 indicating large magnitude classifications (Harris & Rice, 2005). AUC analysis also enables examinations of independent-scale point-estimates, allowing for precise estimates of interpretive accuracy if that scale score is used to make a clinical determination within the sample analyzed. Using a point-estimate approach, we calculated sensitivity (i.e., a test's detection of individuals with a given condition), specificity (i.e., a test's detection rate for those who do not have a given condition), positive

predictive value (PPV; the probability an individual with a test score at/above a given threshold has a particularly condition), and negative predictive value (NPV; the probability an individual with a test score below a given threshold has a particular condition) for the CBS and CB-SOS scales. We used .20, .30, and .40 base rate estimates for malingering during these point estimate calculations because these rates approximate frequent estimates of malingering (Denning & Shura, 2019). We also evaluate incremental predictive utility of the CB-SOS scales beyond CBS using hierarchical logistic regression predicting failure on concurrently administered PVTs. To determine incremental utility, CBS was entered into Block 1 and a single CB-SOS scale was entered into Block 2. We evaluated $F^2\Delta$ to determine the statistical significance in change and $R^2\Delta$ for the magnitude of that effect.

Results

Descriptive information, along with comparisons of scores for the CBS/CB-SOS scales and the PAI scales comprising those scales, are provided in Table 1. Independent sample t-tests indicate significant, medium effect differences for the CBS across the criterion groups for those who passed and failed PVTs: t(286) = 5.342, $p < .001$, d = .70, CB-SOS-1 t(286) = 5.996, $p < .001$, d = .79, CB-SOS-2 t(286) = 5.817, $p < .001$, d = .76, and CB-SOS-3, t(286) = 5.982, $p < .001$, d = .77. Receiver Operator Curve (ROC) analyses on each CB-SOS scale produced large AUC effects (.70 to .71).[1] Using point-estimates created during ROC analysis, sensitivity, specificity, and PPV/NPV at 20%, 30%, and 40% base rates for CB-SOS scales were calculated. Point estimates for CBS and CB-SOS scales are reported in Table 2. Cut scores were identified using the commonly accepted threshold for specificity of .90 and resulting sensitivities were generally low. Setting the specificity to ≥.90, the following cut points were identified: T ≥ 67 for CB-SOS-1, score ≥4.7 for CB-SOS-2, and T ≥ 67 for CBS.

Hierarchical logistic regressions predicting the PVT testing failure demonstrated that only CB-SOS-2 provided incremental identification of groups beyond CBS (i.e., each CB-SOS scale was independently entered into block 1 and CBS was entered into block 2 of each regression, with model change statistics evaluating incremental utility): CB-SOS-1 (F(2,256) = 13.343, $R^2 = .10$, $R^2\Delta < .01$, $F^\Delta = ns$), CB-SOS-2 (F (2,256) = 13.391, $R^2 = .13$, $R^2\Delta = .015$, $F^\Delta = .03$), and CB-SOS-3 (F(2,256) = 15.204, $R^2 = .11$, $R^2\Delta = .01$, $F^\Delta = ns$). Conversely, CBS was incremental (i.e., each CB-SOS scale was independently entered into a model with CBS in model 2) beyond CB-SOS-1 (F

Table 1. Descriptive statistics of CBS, CBS-SOS scales, and component Personality Assessment Inventory Scales.

	Full Sample		Pass All		Fail Any				Correlations			
Scale	M	SD	M	SD	M	SD	d (%95 CI)	t(df = 286)	CBS	CB-SOS-1	CB-SOS-2	CB-SOS-3
CBS	9.8	4.4	9.0	3.9	11.9	4.7	.70 (−.963, −.436)	−5.342**	-	-	-	-
CB-SOS-1	55.8	9.6	53.8	8.5	60.9	10.2	.79 (−1.050, −.520)	−5.996**	0.81	-	-	-
CB-SOS-2	4.0	0.7	3.9	0.7	4.4	0.8	.76 (−1.026, −.497)	−5.817**	0.74	0.92	-	-
CB-SOS-3	57.2	9.3	55.3	8.4	62.2	9.8	.78 (−1.048, −.519)	−5.982**	0.77	0.96	0.96	-
PAI												
NIM	53.2	10.9	51.0	9.0	58.9	13.2	.76 (−1.025, −.497)	−5.811**	0.73	0.85	0.80	0.82
SOM	58.9	11.4	57.4	11.0	62.8	11.6	.48 (−.740, −.220)	−3.664**	0.56	0.73	0.89	0.81
DEP	60.2	13.8	57.5	12.4	66.9	14.9	.72 (−.979, −.452)	−5.465**	0.75	0.91	0.81	0.87
ANX	55.9	12.7	53.3	12.0	62.6	12.1	.77 (−1.038, −.508)	−5.902**	0.68	0.85	0.83	0.80
SCZ	58.5	12.6	56.0	11.4	64.8	13.3	.74 (−.999, −.471)	−5.615**	0.71	0.88	0.74	0.85
SUI	48.3	9.4	47.8	8.7	49.6	11.2	.19 (−.445, .070)	−1.433**	0.47	0.58	0.31	0.40
SOM-C	58.4	13.0	56.7	11.9	62.6	14.8	.46 (−.716, −.196)	−3.482ns	0.55	0.68	0.81	0.77
SOM-S	59.2	11.3	57.5	11.3	63.3	10.5	.53 (−.785, −.264)	−4.009**	0.55	0.67	0.81	0.75
ANX-P	56.2	11.8	54.1	11.0	61.4	12.3	.64 (−.902, −.378)	−4.885**	0.69	0.78	0.80	0.81
DEP-P	62.5	12.4	60.2	11.6	68.4	12.4	.69 (−.949, −.423)	−5.238**	0.57	0.69	0.69	0.74
PAR-R	52.6	12.1	51.3	11.5	55.8	13.0	.37 (−.633, −.115)	−2.855**	0.40	0.64	0.54	0.67

M = Mean, SD = Standard Deviation, d = Cohen's d effect size, %95 CI = 95th Percent Confidence Interval, t/df = independent samples t-test with equal variances assumed and associated degrees of freedom. * = $p < .05$, ** = $p < .01$, ns = non-significant. NIM = Negative Impression Management, SOM = Somatic Complaints, DEP = Depression, ANX = Anxiety, SCZ = Schizophrenia, SUI = Suicidality, SOM-C = Somatic-Conversion, SOM-S = Somatic-Somatization, ANX-P = Anxiety-Physiological, DEP-P = Depression-physiological, PAR-R = Paranoia-Resentment. Calculation of CBS is drawn from Gaasedelen et al. (2019): Item33 [ANX-P] + Item77 [PAR-R] + Item113 [ANX-P] + Item166 [DEP-Affective] + Item206 [DEP-Affective] + Item209 [NIM] + Item242(reversed) [Treatment Rejection; RXR] + Item252(reversed) [SOM-Health Concerns] + Item274 [Anxiety-Related Disorders-Trauma; ARD-T] + Item304(reversed) [Positive Impression Management]. Two versions of the CB-SOS-2 are presented given scoring discrepancy, allowing for a testing of generalization in future CB-SOS studies. Because of the same underlying constructs making up both CB-SOS-2 and CB-SOS-2-AVG, PAI correlations are the same and, as such, are presented across both measures in single values. These two versions reflect the sum of weighted scale and the average sum of each weighted scale (CB-SOS-2 and CB-SOS-2-AVG, respectively). CB-SOS scales were calculated based on formulas outlined by Boress et al. (2021): CB-SOS-1 = (NIM + SOM + DEP + ANX + SCZ + SUI)/6, CB-SOS-2 = (NIM*.015246) + (SOM*.033504)+(ANX*.017804)+(DEP*.010947) + (SCZ*-.002386) + (SUI*-.006888), CB-SOS-2-AVG = [(NIM*.015246) + (SOM*.033504) +(ANX*.017804)+(DEP*.010947) + (SCZ*-.002386) + (SUI*-.006888)]/6, and CB-SOS-3 = (NIM + SCZ + SOM-C + SOM-S + DEP-P + ANX-P + PAR-R)/7.

(2,256) = 13.343, R^2 = .10, $R^{2\Delta}$ = .04, F^Δ(1,254) = 10.933, $p\Delta$ = .001) and CB-SOS-2 (F(2,256) = 13.391, R^2 = .14, $R^{2\Delta}$ = .04, F^Δ(1,254) = 9.434, $p\Delta$ < .01), but not CB-SOS-3 (F(2,256) = 15.204, R^2 = .10, $R^{2\Delta=}$ .01, F^Δ(1,254) = 1.725, $p\Delta$ = ns).

Finally, exploratory analysis was also conducted to examine the potential impact of NIM scores within CB-SOS calculation. NIM, as a validity scale, may be used to predict invalid responding; however, research also shows that elevations on NIM are associated with clinical pathology (see, Bellet et al., 2018). Only 3 participants had NIM>91; when these cases were removed and analyses were re-ran, findings were not substantially different. Thus, NIM's inclusion in the CB-SOS scales does not appear to impact the utility of the scale in discriminating between PVT-based criterion groups.

Discussion

Among military service members there is an elevated risk of brain injury, with incidence rates between 15% and 23% of those serving in Operation Iraqi Freedom (OIF), Operation Enduring Freedom (OEF), and Operation New Dawn (OND; Kong et al., 2022). Along with the acute symptoms associated with brain injury, there is concern for chronic symptomatology among a subset of these individuals (e.g., Lotan et al., 2018).

However, juxtaposed with this concern are the elevated base rates of validity test failure (Armistead-Jehle & Buican, 2012; Armistead-Jehle et al., 2018; Denning & Shura, 2019). Thus, it is unsurprising that there is a major international focus on the development of effective performance-based diagnostic tests (Robinson-Freeman et al., 2020) and that neuropsychologists frequently incorporate broadband personality tests into evaluations for the purpose of symptom validity determinations (Russo, 2018). Developing assessments of brain injury, as well as the associated alterations in affective regulation and impulse control, require an ability to effectively discriminate valid from invalid responding. It also highlights the need for integration of PVT data into PAI assessment research (Fokas & Brovko, 2020). Identifying over-reporting symptoms related to brain injury has long attracted the attention of neuropsychologists working in military settings (e.g., Cooper et al., 2011) and the integration of scales assessing this response style into a popular instrument (which can also provide treatment and diagnostic considerations) increases the utility of standardizing such a practice. Building on this need for better assessment of cognitive symptoms and the detection of invalid responding, this study cross-validated the recently published PAI CB-SOS over-reporting scales (Boress et al., 2021; Shura et al., 2022) in a sample of active-duty Army

Table 2. Classification statistics for the CBS and CB-SOS scales.

	Cut-Score	Sensitivity	Specificity	20% Base Rate			30% Base Rate			40% Base Rate		
				Hit Rate	PPP	NPP	Hit Rate	PPP	NPP	Hit Rate	PPP	NPP
CBS	21	.05	.99	.80	.56	.81	.71	.69	.71	.61	.77	.61
	20	.06	.99	.80	.61	.81	.71	.73	.71	.62	.81	.61
	19	**.07**	**.97**	**.79**	**.35**	**.81**	**.70**	**.48**	**.71**	**.61**	**.59**	**.61**
	18	.11	.96	.79	.39	.81	.70	.52	.72	.62	.63	.62
	17	.17	.95	.80	.47	.82	.72	.61	.73	.64	.70	.63
	16	**.22**	**.93**	**.79**	**.43**	**.83**	**.72**	**.57**	**.74**	**.65**	**.67**	**.64**
	15	.31	.88	.77	.40	.84	.71	.53	.75	.65	.64	.66
CB-SOS-1	83	.02	.99	.80	.39	.80	.70	.52	.70	.60	.63	.60
	78	**.07**	**.99**	**.81**	**.66**	**.81**	**.72**	**.77**	**.71**	**.62**	**.84**	**.62**
	74	.14	.97	.80	.54	.82	.72	.67	.72	.64	.76	.63
	73	.15	.96	.80	.49	.82	.72	.62	.72	.64	.72	.63
	72	.15	.96	.79	.46	.82	.71	.59	.72	.63	.69	.63
	71	.16	.94	.78	.39	.82	.70	.52	.72	.63	.63	.63
	67	.28	.92	.79	.46	.84	.73	.60	.75	.66	.70	.66
	66	.30	.91	.79	.45	.84	.72	.58	.75	.66	.68	.66
	65	.32	.89	.78	.43	.84	.72	.56	.75	.66	.67	.66
CB-SOS-2	5.4	.10	.96	.79	.39	.81	.70	.52	.71	.62	.63	.62
	5.3	**.14**	**.96**	**.79**	**.44**	**.82**	**.71**	**.57**	**.72**	**.63**	**.68**	**.62**
	5.2	.14	.95	.79	.41	.81	.71	.55	.72	.63	.65	.62
	5.0	.22	.94	.80	.49	.83	.73	.62	.74	.65	.72	.64
	4.7	.32	.91	.79	.47	.84	.73	.60	.76	.67	.70	.67
	4.6	.37	.89	.79	.45	.85	.73	.59	.77	.68	.69	.68
	4.5	.40	.86	.77	.41	.85	.72	.55	.77	.67	.65	.68
CB-SOS-3	85	.02	1.00	.80	.56	.80	.70	.69	.70	.61	.77	.60
	79	.06	.99	.80	.61	.81	.71	.73	.71	.62	.81	.61
	76	.09	.98	.80	.47	.81	.71	.61	.71	.62	.70	.62
	75	.12	.98	.81	.56	.82	.72	.69	.72	.63	.77	.63
	74	**.12**	**.97**	**.80**	**.52**	**.82**	**.72**	**.65**	**.72**	**.63**	**.74**	**.62**
	73	.16	.97	.81	.54	.82	.72	.67	.73	.64	.76	.63
	72	.17	.95	.80	.47	.82	.72	.61	.73	.64	.70	.63
	68	.25	.93	.79	.46	.83	.72	.59	.74	.66	.69	.65
	67	.31	.91	.79	.47	.84	.73	.60	.75	.67	.70	.66
	66	.33	.90	.79	.45	.84	.73	.58	.76	.67	.69	.67
	65	.38	.86	.76	.41	.85	.72	.54	.76	.67	.65	.68

Bolded values represent cut-values suggested by Boress et al. (2021) for the CB-SOS scales and for recommended value of CBS in military samples by Armistead-Jehle et al. (2021). We have also bolded the CBS cut score from it's initial validation (CBS ≥ 19; Gaasedelen et al., 2019). PPP = Positive Predictive Power, NPP = Negative Predictive Power.

personnel seen for neuropsychological evaluation. It also directly compared performance of these scales to the already-validated CBS (Armistead-Jehle et al., 2020; Gaasedelen et al., 2019; Shura et al., 2022).

The effects observed in this study approximated moderate differences (Cohen's d = .70 to .79), with large overall classification rates (AUC rounded to .71 for all scales). These findings are commensurate with the initial CB-SOS validation study (Boress et al., 2021) and to the recent study on CB-SOS with Veterans (Shura et al., 2022). However, the results are also distinct as they highlight the need for different cut scores to meet the comparable classification rates. Within active-duty personnel, CB-SOS and CBS perform in a largely similar manner (e.g., comparable sensitivity, specificity, positive and negative predictive power); however, CBS has a small amount of incremental, predictive utility suggesting that it may be the front-line scale. However, calculation of the CBS requires access to PAI item responses and is somewhat more cumbersome to acquire. When those are not available, the CB-SOS scales seem to

represent good alternatives to assess cognitive symptom over-reporting.[2] At optimal cut scores, the CBS and CB-SOS scales produce high specificity, but lower sensitivity within Active-Duty personnel.

Our cross-validation of the CB-SOS scales suggests that the prior cut-scores identified for the CB-SOS scales may not generalize to active-duty personnel. At least in part, performance differences are likely the result of underlying substantive scale (e.g., NIM, SOM, and DEP) elevation patterns on the PAI (see, Gaines et al., 2013 for discussion of this pattern in invalid responding detection more generally on the PAI). More specifically, our PAI scale means were lower than those observed in the Boress et al. (2021) study. As a result of differences in underlying item endorsement/scale elevation, cut-scores needed to achieve a specificity of ≥.90 (Sweet et al., 2021) differed in this study across the CB-SOS scales from those scores recommended during validation (CB-SOS-1 = T78, CB-SOS-2 = 5.3, CB-SOS-3 = T74; Boress et al., 2021). Our results suggest that a slightly lower cut score for CB-SOS-3 (T = 67) and CB-SOS-1 (T = 66) better

achieve desired levels of specificity in the current military sample. CB-SOS-2 also has a lower cut-score, potentially due to the reliance of sample-dependent metrics (regression beta weights) to compute. Importantly, each of the CB-SOS scales produces low sensitivities when specificity is set at ≥.90, generally approximating the lowest observed sensitivity value from initial validation (i.e., sensitivity = .29 [CB-SOS-1] to .41 [CB-SOS-2]; Boress et al., 2021). As such, CB-SOS and CBS should be considered as methods to "rule in" rather than "rule out" the potential of over-reported cognitive symptoms. The CB-SOS scales utilize scale-mean scores in their calculation, while the CBS employs an item-level computed scale. Thus, the CB-SOS scales potentially offer a more accessible method of detecting invalid cognitive responding if performance relative to the CBS is equivalent or superior. While the CBS provided a small degree of improved statistical prediction, our analyses essentially suggest that the CBS and CB-SOS scales are generally equivalent in this regard given the small change in predictive power. Current data then suggest that any of the scales will serve well as the front-line PAI measure of cognitive over-reporting, at least within the military population.

The findings observed in this study also have concrete implications for how CBS and CB-SOS may be integrated into neuropsychological care for Active-Duty personnel. Clinicians working with individuals with high rates of brain injury or cognitive complaints (e.g., military neuropsychological clinics) should use the cut-scores identified in this study, so long as clinical pathology (e.g., depression, anxiety, etc.) is of a similar severity. Conversely, clinicians assessing active-duty personnel who see patterns of pathology on the PAI substantive scales, which differ from those reported in this study should be wary that classification effectiveness across CBS/CB-SOS cut scores is likely to vary. Given that this cross-validation excluded individuals undergoing medical boarding and temporary disability, it is likely that those populations differ in endorsed symptom severity and failed validity testing (Armistead-Jehle et al., 2018). Given the increased rate of disability retirement despite the decrease in disability rates (Gubata et al., 2014), validation in those specific groups is an important area of future study. Likewise, variations in service experience may produce different underlying patterns of medical board evaluation (Thomas et al., 2017), requiring further study.

Limitations and future directions

Several limitations are observed in this investigation. First, the findings may reflect the accuracy of the PVTs used to create the outcome groups across and differ from the initial validation (e.g., MSVT/NV-MSVT versus Test of Memory Malingering [TOMM], Dot Counting Task, or embedded PVTs; Boress et al., 2021). Next, our sample was also predominantly white and male, limiting generalizability beyond those individuals. These limitations notwithstanding, the PAI has several scales that aid in the detection of invalid responding, most of which only undergo an initial validation study and no further psychometric investigations (McCredie & Morey, 2018). This study expanded research on new scales (i.e., CBS and CB-SOS) which assess cognitive over-reporting, supplementing a needed area of research for the PAI. It also provides direct guidance for clinicians working with military populations. Given the findings that scale performance did not generalize between this study and other samples, further replication on the PAI with other neuropsychological populations and with additional PVT criterion classification methods is warranted (see, Fokas & Brovko, 2020). Evaluation of effectiveness in medical board and temporary disability evaluations is also needed. Cognitive over-reporting scales (e.g., CB-SOS) are not considered best practice for the detection of primarily feigned or exaggerated psychiatric symptoms (Sherman et al., 2020). However, research remains warranted on exaggerated emotional responses on CBS and CB-SOS given the relatively similar effect sizes across different types of embedded broadband over-reporting indicators (e.g., Hawes & Boccaccini, 2009) and the potential impact of psychiatric symptom level on effectiveness of the scales.

Note

1. AUC and classification accuracies range from 0 (completely inaccurate classification) to 1.00 (completely accurate classification), with a value of .50 indicating classification at random chance levels. AUC values were interpreted as having small (.57), medium (.64), and large (.71) effects sizes (Rice & Harris, 2005).
2. No support was received for this project from PAR, Inc, publisher and distributor of the PAI.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

ORCID

Paul B. Ingram 🔟 http://orcid.org/0000-0002-5409-4896

Patrick Armistead-Jehle 🆔 http://orcid.org/0000-0002-6784-1432
Tristan T. Herring 🆔 http://orcid.org/0000-0001-7468-794X
Nicole M. Morris 🆔 http://orcid.org/0000-0001-6565-9816

Data availability statement

References

Armistead-Jehle, P., & Buican, B. (2012). Evaluation context and symptom validity test performance in a U.S. Military sample. *Archives of Clinical Neuropsychology*, *27*(8), 828–839. https://doi.org/10.1093/arclin/acs086

Armistead-Jehle, P., Cole, W. R., & Stegman, R. L. (2018). Performance and symptom validity testing as a function of medical board evaluation in US military service members with a history of mild traumatic brain injury. *Archives of Clinical Neuropsychology*, *33*(1), 120–124. https://doi.org/10.1093/arclin/acx031

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*, *27*(1), 101–113. https://doi.org/10.1093/arclin/acr081

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. (2012). Addressing challenges to MMPI-2-RF-based testimony: Questions and answers. *Archives of Clinical Neuropsychology*, *27*(7), 691–705. https://doi.org/10.1093/arclin/acs083

Boress, K., Gaasedelen, O. J., Croghan, A., Johnson, M. K., Caraher, K., Basso, M. R., & Whiteside, D. M. (2021). Validation of the Personality Assessment Inventory (PAI) scale of scales in a mixed clinical sample. *The Clinical Neuropsychologist*, 1–16. https://doi.org/10.1080/13854046.2021.1900400

Burchett, D., & Bagby, R. M. (2021). Assessing negative response bias: A review of the noncredible overreporting scales of the MMPI-2-RF and MMPI-3. *Psychological Injury and Law*, *14*(1), 1–15. https://doi.org/10.1007/s12207-021-09405-1

Bush, S. S., Ruff, R. M., Troster, 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 and Planning Committee. *Archives of Clinical Neuropsychology*, *20*(4), 419–426. https://doi.org/10.1016/j.acn.2005.02.002

Carone, D. A. (2009). Test review of the medical symptom validity test. *Applied Neuropsychology*, *16*(4), 309–311. https://doi.org/10.1080/09084280903297883

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Revised ed.). Academic Press.

Cooper, D. B., Nelson, L., Armistead-Jehle, P., & Bowles, A. O. (2011). Utility of the mild brain injury atypical symptoms scale as a screening measure for symptom over-reporting in operation enduring freedom/operation Iraqi Freedom service members with post-concussive complaints. *Archives of Clinical Neuropsychology*, *26*(8), 718–727. https://doi.org/10.1093/arclin/acr070

Denning, J. H., & Shura, R. D. (2019). Costs 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

Department of Veteran Affairs. (2022). *VA/DOD clinical practice guidelines*. https://www.healthquality.va.gov/

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

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

Gaines, M. V., Giles, C. L., & Morgan, R. D. (2013). The detection of feigning using multiple PAI scale elevations: A new index. *Assessment*, *20*(4), 437–447. https://doi.org/10.1177/1073191112458146

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a response bias scale for the MMPI-2. *Assessment*, *14*(2), 196–208. https://doi.org/10.1177/1073191106295861

Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows: User's manual*. Green's Publishing

Green, P. (2008). *Manual for the nonverbal medical symptom validity test*. Green's Publishing.

Gubata, M. E., Packnett, E. R., & Cowan, D. N. (2014). Temporal trends in disability evaluation and retirement in the Army, Navy, and Marine Corps: 2005–2011. *Disability and Health Journal*, *7*(1), 70–77. https://doi.org/10.1016/j.dhjo.2013.08.003

Harwood, T. M., Beutler, L. E., & Groth-Marnat, G. (2011). *Integrative assessment of adult personality*. Guilford Press.

Hawes, S. W., & Boccaccini, M. T. (2009). Detection of over-reporting of psychopathology on the personality assessment inventory: A meta-analytic review. *Psychological Assessment*, *21*(1), 112. https://doi.org/10.1037/a0015036

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. L., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*(7), 1093–1129. https://doi.org/10.1080/13854040903155063

Ingram, P. B., Schmidt, A. T., Bergquist, B., & Currin, J. M. (2022). Education, instrument exposure, and competence in psychological assessment: A national survey of practices and beliefs of health service psychology trainees. *Training and Education in Professional Psychology*, 16(1), 10–19. https://doi.org/10.1037/tep0000348

Ingram, P. B., Tarescavage, A. M., Ben-Porath, Y. S., & Oehlert, M. E. (2020). The MMPI-2-restructured form (MMPI-2-RF) validity scales: Patterns observed across veteran affairs settings [Measurement-based care and psychological assessment in mental health services]. *Psychological Services*, 17(3), 355–362. https://doi.org/10.1037/ser0000339

Keiski, M. A. (2007). *Use of the Personality Assessment Inventory (PAI) following traumatic brain injury* [Doctoral Dissertation]. University of Windsor. ProQuest Dissertations Publishing. https://scholar.uwindor.ca/etd/4710

Kong, L. Z., Zhang, R. L., Hu, S. H., & Lai, J. B. (2022). Military traumatic brain injury: A challenge straddling neurology and psychiatry. *Military Medical Research*, 9(1), 1–18. https://doi.org/10.1186/s40779-021-00363-y

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., ... Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454. https://doi.org/10.1037/abn0000258

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18(4), 1–7. https://doi.org/10.1017/S1355617712000240

Lotan, E., Morley, C., Newman, J., Qian, M., Abu-Amara, D., Marmar, C., & Lui, Y. W. (2018). Prevalence of cerebral microhemorrhage following chronic blast-related mild traumatic brain injury in military service members using susceptibility-weighted MRI. *American Journal of Neuroradiology*, 39(7), 1222–1225. https://doi.org/10.3174/ajnr.A5688

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. https://doi.org/10.1037/pas0000574

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 sample. *Journal of Military Psychology*. https://doi.org/10.1080/08995605.2021.2013059

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

Ratcliffe, L. N., Hale, A. C., Gradwohl, B. D., & Spencer, R. J. (2022). *Preliminary findings from reevaluating the MMPI Response Bias Scale items in veterans undergoing neuropsychological evaluation* (pp. 1–8). Adult. https://doi.org/10.1080/23279095.2022.2106571

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

Robinson-Freeman, K. E., Collins, K. L., Garber, B., Terblanche, R., Risling, M., Vermetten, E., Besemann, M., Mistlin, A., & Tsao, J. W. (2020). A decade of mTBI experience: What have we learned? A summary of proceedings from a NATO lecture series on military mTBI. *Frontiers in Neurology*, 11, 836. https://doi.org/10.3389/fneur.2020.00836

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

Rosnow, R. L., Rosenthal, R., & Rubin, D. B. (2000). Contrasts and correlations in effect-size estimation. *Psychological Science*, 11(6), 446–453. https://doi.org/10.1111/1467-9280.00287

Russo, A. C. (2018). A practitioner survey of Department of Veterans Affairs psychologists who provide neuropsychological assessments. *Archives of Clinical Neuropsychology*, 33(8), 1046–1059. https://doi.org/10.1093/arclin/acx139

Sharf, A. J., Rogers, R., Williams, M. M., & Henry, S. A. (2017). The effectiveness of the MMPI-2-RF in detecting feigned mental disorders and cognitive deficits: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment*, 39(3), 441–455.

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*, 35(6), 735–764. https://doi.org/10.1093/arclin/acaa019

Shura, R., Ingram, P. B., Miskey, H. M., Martindale, S. L., Rowland, J. A., & Armistead-Jehle, P. (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*. https://www.tandfonline.com/doi/abs/10.1080/13854046.2022.2131630

Slavin-Mulford, J., Sinclair, S. J., Stein, M., Malone, J., Bello, I., & Blais, M. A. (2012). External validity of the Personality Assessment Inventory (PAI) in a clinical sample. *Journal of Personality Assessment*, 94(6), 593–600. https://doi.org/10.1080/00223891.2012.681817

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., & Boone, K. B., Conference Participants. (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*, 1–54. https://www.tandfonline.com/doi/full/10.1080/13854046.2021.1896036

Thomas, W. A., Jr, Doane, E. L., Gallavan, R. H., Tavares, S., & Jones, M. C. (2017). Medical evaluation board for mental health condition: US Army officer medical evaluation board data by branch and component. *Military Medicine*, 182(9–10), e1908–e1916. https://doi.org/10.7205/MILMED-D-17-00041

Till, C., Christensen, B. K., & Green, R. E. (2009). Use of the Personality Assessment Inventory (PAI) in individuals with traumatic brain injury. *Brain Injury*, *23*(7–8), 655–665. https://doi.org/10.1080/02699050902970794

Van Voorhees, E. E., Dennis, P. A., Elbogen, E. B., Clancy, C. P., Hertzberg, M. A., Beckham, J. C., & Calhoun, P. S. (2014). Personality assessment inventory internalizing and externalizing structure in veterans with posttraumatic stress disorder: Associations with aggression. *Aggressive Behavior*, *40*(6), 582–592. https://doi.org/10.1002/ab.21554

Wager, J. G., & Howe, L. L. S. (2010). Nonverbal medical symptom validity test: Try faking now! *Applied Neuropsychology*, *17*(4), 305–309. https://doi.org/10.1080/09084282.2010.525093

Whiteside, D. M., Wald, D., & Busse, M. (2011). Classification accuracy of multiple visual spatial measures in the detection of suspect effort. *The Clinical Neuropsychologist*, *25*(2), 287–301. https://doi.org/10.1080/13854046.2010.538436

Wright, C. V., Beattie, S. G., Galper, D. I., Church, A. S., Bufka, L. F., Brabender, V. M., & Smith, B. L. (2016). Assessment practices of professional psychologists: Results of a national survey. *Professional Psychology: Research and Practice*, *48*(2), 73–78. https://doi.org/10.1047/pro000008

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. https://doi.org/10.1080/13854046.2016.1159730