Psychological Assessment

Interpreting the Personality Assessment Inventory (PAI) Validity Scales: Leveraging Population-Level Veteran Affairs (VA) Data From 2008 to 2024

Robert D. Shura, Paul B. Ingram, Ryan W. Schroeder, and Patrick Armistead-Jehle

Online First Publication, July 28, 2025. https://dx.doi.org/10.1037/pas0001403

CITATION

Shura, R. D., Ingram, P. B., Schroeder, R. W., & Armistead-Jehle, P. (2025). Interpreting the Personality Assessment Inventory (PAI) validity scales: Leveraging population-level Veteran Affairs (VA) data from 2008 to 2024. *Psychological Assessment*. Advance online publication. https://dx.doi.org/10.1037/pas0001403

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

In the public domain
ISSN: 1040-3590

https://doi.org/10.1037/pas0001403

Interpreting the Personality Assessment Inventory (PAI) Validity Scales: Leveraging Population-Level Veteran Affairs (VA) Data From 2008 to 2024

Robert D. Shura[1, 2, 3], Paul B. Ingram[4], Ryan W. Schroeder[5], and Patrick Armistead-Jehle[6]

[1] W. G. (Bill) Hefner VA Healthcare System, Salisbury, North Carolina, United States
[2] VA Mid-Atlantic Mental Illness Research, Education, and Clinical Center, Durham, North Carolina, United States
[3] Department of Neurology, Wake Forest University School of Medicine
[4] Department of Psychological Sciences, Texas Tech University
[5] Robert J. Dole VA Medical Center, Wichita, Kansas, United States
[6] Munson Army Health Center, Fort Leavenworth, Kansas, United States

The purpose of this study was to examine base rates of invalidity and intercorrelations of validity scales on the Personality Assessment Inventory (PAI) in a population of U.S. veterans. All PAIs administered in the Veteran Affairs ($N = 36,830$) using mental health assistant through 03/2024 were pulled from the Veteran Affairs Corporate Data Warehouse. Basic demographic data and stop codes (codes identifying the type of clinic the measure was administered in) were also pulled. Base rates of invalidity on all known validity scales ($N = 18$) were calculated in general, cumulatively, and by stop code. Spearman's correlations were run to compare all validity scales to one another. Broadly, base rates were highest but variable for overreport scales (10.1%–29.9%), with underreport (0.4%–7.5%) and noncontent response (3.4%–4.5%) rates far lower. Large correlations were demonstrated among the various overreport scales, with the exception of Rogers Discriminant Function. Underreport scales were also highly correlated with each other, with the exception of Cashel's Discriminant Function. Elevations on noncontent response scales were uncommon across the sample. Underreport scales were also rare, save in the context of presurgical evaluations. Overreporting was highest in the disability clinics, lower across clinical assessment clinics, and lowest in the presurgical assessment clinic. Invalidity base rate patterns from this study were generally consistent with a prior study using the Minnesota Multiphasic Personality Inventory-2-Restructured Form. Results highlight the importance of symptom validity assessment with veterans, most notable for overreporting in disability claims and underreporting in presurgical evaluations.

Public Significance Statement
This study examines veterans' endorsement patterns on the Personality Assessment Inventory (PAI) validity scales, utilizing every PAI administered within the Veteran Administration health care system and stored within the Corporate Data Warehouse ($N = 36,830$). Results indicate that veterans commonly score beyond established cutoffs on validity scales, especially those assessing exaggeration or over-endorsement of symptoms.

Keywords: Personality Assessment Inventory, veteran, base rates, response bias, validity

The Veteran Affairs (VA) is the nation's largest health care organization, caring for over 18 million veterans (Department of Veteran Affairs, 2024), nearly a quarter of whom receive services for conditions related to their service (Kessler et al., 2014). The VA provides quality care (Apaydin et al., 2023; Asch et al., 2004), consistent with the Veterans Health Administration's longstanding investment in training and developing health care advancements. This is a necessary endeavor, as health care needs of veterans are uniquely complex and multifaceted. Veterans are subject to unique experiences stemming from their service and deployments (e.g.,

Julie Suhr served as action editor.

Robert D. Shura https://orcid.org/0000-0002-9505-0080

The views, opinions, and/or findings contained in this article are those of the authors and should not be construed as an official Veteran Affairs (VA), Department of Defense, or U.S. Government position, policy, or decision unless so designated by other official documentation. The authors declare no conflicts of interest, financial, or otherwise.

This work was supported by the W. G. (Bill) Hefner VA Health Care System and the VA Mid-Atlantic (VISN 6) Mental Illness Research, Education, and Clinical Center. The material is also the result of work supported with resources and the use of facilities at Robert J. Dole VA Medical Center. There was no other funding supporting this study.

Robert D. Shura played a lead role in conceptualization, data curation, formal analysis, investigation, project administration, and software and an equal role in methodology, resources, visualization, writing–original draft, and writing–

continued

Gulf War Syndrome, Agent Orange exposure), each of which has unique physiological and psychological effects (Haley et al., 1997). Veterans also experience disproportionately high rates of psychiatric and medical conditions, regularly having worse outcomes and greater impairments compared to the general population (Kilpatrick et al., 2013; Trivedi et al., 2015). The VA's interwoven forensic context (Russo, 2014, 2018; Worthen & Moering, 2011), the distinctive life experiences (Glenn et al., 2002; Ingram et al., 2020a, 2020b; MacLean & Elder, 2007), demographic factors (Denning & Horner, 2024) of veterans, and military culture broadly (Kümmel, 2018) work to create novel health care presentations and needs. Consequently, diagnosing and treating veterans requires tailored and often highly nuanced approaches, extending to clinics outside of the VA where many services are offered (West & Weeks, 2009).

One of the major barriers to effective care for veterans is the inadequacy of existing assessment practices to fully capture the complexity and high prevalence of pathology (e.g., Williams et al., 2022). This gap in assessment tools and methodologies underscores a need for innovation. Given the substantive infrastructure and data warehousing that is available to researchers, the VA is uniquely positioned to leverage big data to help solve these problems (Fihn et al., 2014; Nelson et al., 2014). For instance, several articles have used the Corporate Data Warehouse (CDW) database to provide contextual information for one commonly used personality assessment instrument (i.e., Minnesota Multiphasic Personality Inventory-2-Restructured Form [MMPI-2-RF]; Ingram et al., 2020a, 2020b). The current article uses the same methodology described by Ingram and colleagues (e.g., extracting data from the CDW, where all medical records are stored, as accessed by the VA Informatics and Computing Infrastructure) to provide the first such study with the Personality Assessment Inventory (PAI).

The Personality Assessment Inventory

The PAI (Morey, 2007, 2020) is a commonly used measure of personality and psychopathology, often employed in clinical practice (Archer et al., 2006). It is widespread acceptance (Kurtz & Blais, 2007; Kurtz & Pintarelli, 2024) and utility (Blais et al., 2010) extend to the Department of Veteran Affairs. The PAI has a substantive research base tying it to veterans and their needs, far beyond its available comparison sample (Morey, 2007). For instance, the PAI has been studied as an instrument that can improve diagnostic accuracy of, and conceptualizations for, different highly prevalent conditions, including posttraumatic stress disorder (PTSD; Bellet et al., 2018; Calhoun et al., 2010; Ingram et al., 2022; Mozley et al., 2005), brain injury (Kennedy et al., 2015; Miskey et al., 2015; Velikonja et al., 2010), and other pathologies and clinical needs (violence risk, suicidality, substance use, personality disorders, etc.; Blalock et al., 2020; Breshears et al., 2010; Crawford et al., 2007; Ruchensky et al., 2024; Shura, Miskey, et al., 2023). The PAI has also been validated for use in treatment prediction (Cersosimo et al., 2022; McCredie et al., 2018), although studies in this area are less common with veterans relative to other populations.

Importantly, the PAI contains numerous validity scales that speak to the reliability of the respondent data and can thus work to better assure accuracy of the resulting clinical interpretations and recommendations (Kurtz & Pintarelli, 2024; Meaux et al., 2022). Overreporting has been identified as the most (Williams et al., 2022; Wygant et al., 2010) frequent pattern of protocol invalidity seen in the VA (Braxton et al., 2007), but this varies greatly between evaluative setting and VA service clinic (Ingram et al., 2020b). The PAI contains a variety of scales designed and validated for the assessment of overreporting, and these scales are frequently used in clinical practice for this purpose (Martin et al., 2025). For comprehensive reviews of each scale and its evidence, readers are referred elsewhere (see Hawes & Boccaccini, 2009; Kurtz & McCredie, 2022; McCredie & Morey, 2018).

PAI overreporting literature includes studies on the primary indicators (i.e., Negative Impression Management [NIM]), as well as secondary (e.g., Multiscale Feigning Index [MFI; Gaines et al., 2013]) and emerging (e.g., Cognitive Bias Scale [CBS; Gaasedelen et al., 2019], Cognitive Bias Scale of Scales [CB-SOS; Boress et al., 2022], Hong Malingering Index [HongM; Hong & Kim, 2001]) scales. Broadly, this research has found that overreporting scales work in a medium to large range of effect across the different detection strategies and content foci (e.g., Boress et al., 2022; Harrison et al., 2022; Morris et al., 2022; Russell & Morey, 2019). Across this PAI literature, effect sizes are typically smaller in criterion-based studies—which have the greatest ecological validity (Rogers & Bender, 2018; Schroeder et al., 2021). Moreover, the consistent and strong performance by the secondary and emerging overreporting scales (e.g., CBS, CB-SOS; Shura, Ingram, et al., 2023; Schroeder et al., 2025) is consistent with calls for expanded study on these scales given their promise (McCredie & Morey, 2018). While most scales have a strong criterion-based, clinically-derived research base, MFI and HongM have seen little replication and cross-validation efforts. The lack of a substantive and clinically derived validation is problematic, as simulation studies overestimate classification accuracy (Rogers & Bender, 2018; Schroeder et al., 2021). The Rogers Discriminant Function (RDF) scale is an exception to the generally strong pattern of PAI overreporting scale performance. RDF regularly produces smaller magnitude effects outside a range that would indicate clinical utility (e.g., Armistead-Jehle & Buican, 2012; Gaines et al., 2013; Harrison et al., 2022) and has performed poorly in criterion groups meta-analyses (Hawes & Boccaccini, 2009). Notably, only a limited number of studies to date have included either veteran (Shura, Ingram, et al., 2023) or military (Armistead-Jehle & Buican, 2012; Morris et al., 2022) samples.

Other forms of response bias (e.g., noncontent invalidity or underreporting) are less prevalent in general and have notably smaller literature bases (and applicable, or validated, scales) for the PAI. The Inconsistency (ICN) and Infrequency (INF) scales for the PAI serve

review and editing. Paul B. Ingram played an equal role in conceptualization, methodology, resources, visualization, writing–original draft, and writing–review and editing. Ryan W. Schroeder played an equal role in conceptualization, methodology, visualization, writing–original draft, and writing–review and editing. Patrick Armistead-Jehle played an equal role in methodology, visualization, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Robert D. Shura, W. G. (Bill) Hefner VA Healthcare System, 1601 Brenner Avenue (11-M), Salisbury, NC 28144, United States. Email: Robert.Shura2@va.gov

as the instrument's primary noncontent invalidity indicators, with the Hong Randomness Index (HongR; Hong & Kim, 2001) as a supplemental indicator. The Positive Impression Management (PIM) scale is the primary underreporting validity index, but there are supplemental scales including the Hong Defensiveness Index (HongD; Hong & Kim, 2001), Defensiveness Index (DEF; Morey, 2007), and Positive Distortion Scale (PDS; Mogge et al., 2017). Although some support for these scales has been demonstrated, many have limited cross-validation or lack comprehensive validation efforts in nonsimulation samples.

The Present Study

Given that detection of invalidity is a critical first step in assessment interpretation of multiscale self-report measures (Burchett & Ben-Porath, 2019), this study examined elevation patterns of the PAI validity scales in a large veteran sample. The sample is based on evaluations from 2008 to 2024, encompassing every PAI administered in the VA electronic medical record during that timeframe. We examined base rates of elevated scores and concurrent relationships between existing scales, including concurrent elevation rates and correlations between constructs measured by the various symptom validity scales. We also evaluated differences as a function of the purpose (e.g., compensation and pension evaluation status, general clinical status, presurgical evaluation status) and evaluation location (i.e., specific VA clinics via stop codes).

Based on prior research with the MMPI-2-RF, and sympton validity test (SVT) research in veteran samples more generally, we developed six main hypotheses. First (H1), most PAI administrations across various VA clinic settings will produce valid protocols based on the core PAI scales. Second (H2), invalid protocols will most frequently be due to overreporting. Third (H3), invalidity rates will be somewhat lower than identified with the MMPI-2-RF within this setting. Fourth (H4), those who invalidate PAI protocols will most likely invalidate the protocol on multiple scales rather than individual ones. These first four hypotheses were based on a study by Ingram et al. (2020b) using the MMPI-2-RF; thus, we expect similar findings with the PAI. Fifth (H5), correlations between overreporting scales will suggest largely overlapping core variance, regardless of the scale's construction or content coverage (Schroeder et al., 2025). And last (H6), when invalid profiles are produced, these will vary by clinic and evaluation setting. Specifically, overreporting elevations will be more common in compensation-seeking evaluation contexts than in general clinical contexts and specialty contexts where underreporting is generally thought to be more likely (presurgical contexts). This pattern was seen on the MMPI-2-RF (Ingram et al., 2020b)

Method

Sample

The VA CDW was used to identify all veterans in the VA system who completed the PAI within the mental health assistant (MHA) and the new web-based MHA Web. MHA is the electronic interface for administrating, scoring, and reviewing results of multiple psychological questionnaires, including the PAI, and it is embedded within the electronic medical record platform. All PAI protocols available in MHA were pulled, leading to a total of 36,830 PAI protocols; the protocols were administered between February 2008 and March 2024. This timeline was not selected by the authors; rather, it reflects when the PAI was added to MHA through the most recent administrations at the time of the data pull. All PAI protocols were administered in either a clinical or compensation and pension context in various clinic types across all VA medical centers and clinics in the United States. Basic demographic information is presented in Table 1. Most of the sample was male (84.9%), White (71.3%), married (44.6%), and receiving some form of service-connected disability (85.9%). The mean age was 46.55 years.

Several demographic factors (age, sex, and ethnicity) were modeled via logistic regression predicting failing any one or more PAI SVT: The model was significant, $\chi^2(1) = 160.16$, $p < .001$, though the effect was small (area under the curve; AUC = .538). Specifically, significance was driven by age ($\chi^2 = -0.01$, $p < .001$) and ethnicity ($\chi^2 = 0.20$, $p < .001$), but not sex ($\chi^2 = 0.03$, $p = .052$). Given the small estimates in the model and near-chance AUC more generally, significant results are likely an artifact of the substantial sample size and without meaningful clinical relevance.

Measures

Personality Assessment Inventory

The PAI is a self-report, multiscale psychological inventory assessing symptom validity, personality constructs, psychopathology, and other areas of clinical interest. The PAI contains 344 items rated on a Likert-like scale across options of *false*, *slightly true*, *mainly true*, and *very true*. The MHA score report and data output in CDW include only the four primary validity scales; thus, all other validity scales were coded by the study authors. Within the MHA system, item options are coded from 1 to 4; however, most scale development uses item coding of 0 to 3; items were recoded as needed when calculating supplemental scales. For all validity indices, cutoff scores are based on manual and published recommendations.

Table 1
Sample Demographics (N = 36,830)

Class	Variable	*M*	*SD*	Range	*N*	%
Sex	Male				31,266	84.9
	Female				5,564	15.1
Race	White				26,263	71.3
	Black				7,334	19.9
	Asian				326	0.9
	Native				615	1.7
	Pacific				383	1.0
Ethnicity	Hispanic				1,893	5.4
	Not Hispanic				33,113	94.6
Marital	Married				16,432	44.6
	Divorced				9,558	26.0
	Separated				1,879	5.1
	Never Married				7,641	20.8
	Single				48	0.1
	Widow/er				659	1.8
	Unknown				613	1.7
SC	Yes				31,650	85.9
	No				5,180	14.1
SC	Total percent	78.33	25.9	0–100		
Age	Years	46.55	14.23	15–114		

Note. SC = service-connected disability.

Primary Validity Scales

The PAI contains four primary validity scales that are profiled with the clinical scales. ICN consists of 10 pairs of similar items, half that should be endorsed differently and half that should be endorsed similarly, thus measuring inconsistent responding, a type of noncontent response bias. INF consists of 12 items that should be infrequently endorsed by the general population and measures careless or random responding; like ICN, this scale is also a noncontent response bias scale. PIM contains 12 items and measures underreporting of psychological distress. NIM includes 12 items rarely endorsed in clinical samples, with items developed to detect exaggerated or feigned symptoms.

Supplementary Validity Indices

Several additional overreport and underreport validity indices are included in the 2007 PAI test manual. On the overreport side, the Malingering Index (MAL) is made up of eight features (scale elevations of scale comparisons) with a total raw score ranging from 0 to 8. A raw score of ≥3 is recommended to identify invalid responding. RDF is a regression-based scale using β weights across 20 different scales and subscales; 12 are weighted in the positive direction, and the remaining in the negative direction. Raw scores > 0 are considered reflective of invalidity. Regarding underreporting scales, DEF is like MAL in that it is calculated from nine different scale features and relationships, each positive one earning a point for a total of 0 to 9. Raw scores of ≥6 indicate invalidity. Cashel's Discriminant Function (CDF), like RDF, is calculated from six scales using β weights, four positive and two of which are negative. Scores of ≥168 are considered invalid due to underreporting.

Additional Published Validity Indices

In addition to the indices in the 2007 PAI manual, several additional scales have since been developed and published (several of which are included in the PAI Plus Manual Supplement [Morey, 2020]). One set of scales created by Hong and Kim (Hong & Kim, 2001) sample all three types of response bias and employ β weights and various scales for raw score calculation. The HongM uses a cutoff of ≥1.59 to identify overreporting, the HongD uses a cutoff of ≥0.92 to identify underreporting, and the HongR uses a cutoff of ≥2.08 to identify random response. The MFI (Gaines et al., 2013) is the average T score of seven clinical scales, with a recommended cutoff of >74 T to identify overreporting. Two additional scales were created from summing additional items. The Negative Distortion Scale (NDS; Mogge et al., 2010) is the sum of 15 items, with a raw score of ≥13 used to identify overreporting. In contrast, the PDS (Mogge et al., 2017) is the sum of 17 items, with a cutoff of ≥45 for identifying underreporting.

Next, several scales have been developed recently to identify the overreport of cognitive symptoms. The first of these was the CBS (Gaasedelen et al., 2019). This measure was designed to predict invalid scores on Performance Validity Tests (PVTs). It consists of 10 PAI items, with an originally identified cut score of ≥19. After CBS' development, three additional indices were created to identify cognitive overreporting using scale elevations (in contrast to the item-level data used to create CBS). These measures are collectively referred to the CB-SOS (Boress et al., 2022). CB-SOS1 is the T score mean of six scales, with a cutoff of ≥78; CB-SOS2 weighted those same scales based on a binary logistic regression to invalid scores on PVTs, with a cutoff score of ≥5.3; and CB-SOS3 is the mean of the seven scales or subscales with the highest AUC, with overreporting identified as ≥74.

Procedure

The study was reviewed by the Salisbury Veterans Affairs Medical Center Institutional Review Board and determined to be exempt. Data were obtained via a Data Access Request Tracker request; data were initially pulled using Microsoft SQL Server. All analyses were conducted using SAS Enterprise Guide. Supplementary scales were calculated using data available in the PAI manuals and published materials. Descriptive and base rate data were calculated for all scales across a variety of cutoff scores. Next, reverse cumulative base rates were calculated for all published noncontent scales, all underreport scales, all overreport scales, and all 18 scales combined. This procedure was repeated using three additional groups of scales: the core invalidity scales, the additional invalidity scales in the 2007 manual, and various scales with empirical support based on published research.

Finally, we evaluated invalidity rates using the most conservative scores for all SVTs based on VA clinic type. VA stop codes are numbers that identify the type of clinic in which a service is rendered. A given clinic can have a primary and a secondary stop code. For example, a typical psychological assessment clinic might have a primary stop code of 510, indicating this is a clinic for individual patients (instead of a group clinic) with services by a psychologist, and a secondary stop code of 538, indicating the clinic is specifically for psychological assessment (as opposed to therapy). Like the criteria used by Ingram and colleagues (Ingram et al., 2020b), we evaluated all primary stop codes with 100 or more PAI administrations, leading to 18 different clinic codes. In addition, encounters in clinics for compensation and disability evaluations (C&P) are identified via nine different secondary stop codes. All administrations in clinics with any of the C&P stop codes were separated into one disability exam category. In addition to calculating invalid rates for all 18 scales, we also calculated the rate for being invalid on any one or more scales for each clinic code.

Transparency and Openness

This study was not preregistered. Data, analytic code, and materials are not available but may be shared on reasonable request within administration guidelines and protocols.

Results

Table 2 presents invalidity rates at varying cutoff scores for the four primary validity indices. In support of H1, 89.1% produced valid profiles based on the primary four indices and skyline cutoff scores. Noncontent responding and underreporting rates were less than 5% across ICN, INF, and PIM, using conservative cutoffs from the PAI manual. In contrast, the rate of overreporting was 10.1%, between two and three times the rate of noncontent response and underreporting, which supports H2. On the MMPI-2-RF, as reported in Ingram et al. (2020b), 57.8% of their over 17,000 veteran sample were fully valid

Table 2
Descriptive and Cumulative Base Rates for Primary Symptom Validity Scales

Scale/≥cutoff	Citation	*M*	*SD*	Min.–Max.	Kurtosis	Skew	Cum.%
ICN (n = 36,830)	Morey (2007)	54.05	9.48	34–112	0.34	0.54	
64							18.1
73							4.5
INF (n = 36,830)	Morey (2007)	52.78	9.42	40–133	1.17	0.93	
60							18.3
75							3.4
NIM (n = 36,830)	Morey (2007)	66.50	17.17	44–144	0.69	0.91	
70							40.3
73							33.5
77							27.1
84							17.2
92							10.1
110							2.3
PIM (n = 36,830)	Morey (2007)	42.09	12.18	15–77	−0.46	0.14	
50							30.0
57							14.3
61							7.5
64							5.3
68							3.5
72							0.6

Note. Grey shaded rows are the invalid cutoffs per manual (Morey, 2007). Min. = minimum; Max. = maximum; Cum. = cumulative; ICN = Inconsistency; INF = Infrequency; NIM = Negative Impression Management; PIM = Positive Impression Management.

on all primary validity indices, thus supporting H3 that rates of invalidity would be lower on the PAI than on the MMPI-2-RF.

Table 3 presents invalidity rates across all PAI supplementary SVTs, which were much more variable. For overreport scales, MAL and HongM were largely consistent with NIM rates at 11% and 9%, respectively; however, the RDF rate was nearly 30% at the manual cutoff. All four of the CBSs as well as the NDS showed rates between 22% and 30%, closer to RDF but much higher than the NIM/MAL/HongM rates around 10%. For underreport supplementary scales, rates were far lower and lower than the PIM rate of 3.5%, except for CDF reaching 7.5% invalid. Finally, the single noncontent supplementary scale, HongR, resulted in an invalid rate like ICN and INF, the three ranging from 3.1% to 4.5%. In sum, invalidity rates across scale type resulted in a pattern of overreport > noncontent response > underreport.

This pattern of invalid responding is demonstrated again when reviewing cumulative rates of invalidity, as shown in Table 4. Considering all PAI validity scales, approximately 10% engaged in significant noncontent response, while less than 4% engaged in any form of underreporting. However, over half the sample showed evidence of overreporting to some degree, with 38.6% producing invalid scores on two or more scales. In contrast, the highest invalidity rate on a single scale was 30.4% (MFI), supporting H4. Combining all three types of validity measures, only 38.4% produced a completely valid profile without evidence of any type of response bias. On examining specific groupings, rates on noncontent and underreport invalidity remain relatively similar, likely due to the fewer total number of scales available. However, greater variability is seen with overreport scales, which seems to be largely driven by the cognitive overreport scales.

Correlations across scales were completed using Spearman's correlations, results of which are presented in Table 5. Collectively, the highest correlations were seen among the overreport scales, though RDF was by far the least related to other overreport scales. Not considering the relationships with RDF (M = .25, SD = .04), correlations between overreporting scales ranged from .54 (CBS and MAL) to .96 (SOS2 to SOS3), which supports the high level of overlapping core variance (H5) across overreport scales (M = .78, SD = .11). Underreport scale correlations were lower than overreport correlations, ranging from -.08 to .73 (PIM to DEF), but also generally high, with the notable outlier of CDF. Noncontent scales were more variable, ranging from .15 (ICN to INF) to .44 (ICN to Hong Randomness), with INF being the outlier for this group of scales.

On examining the clinics where the PAI was administered most frequently, C&P clinics were the highest at over 10,000 administrations, though psychological testing clinics were a close second at 9,581 administrations (Tables 6 and 7). A similar pattern across SVTs is seen in which overreport > noncontent > underreport; however, rates within SVTs varied significantly by clinic. Tables 6 and 7 show invalid rates for all SVTs by stop code. The final column shows invalidity rates based on any one or more invalid measures. The highest rate is seen in the substance abuse group stop code, whereas the lowest was in the mental health biomedical stop code. Three clinics had the highest invalidity rate on three scales: General internal medicine had the highest rate for MAL and HongM, but also on DEF; individual psychiatry had the highest rates for NIM, SOS1, and NDS (all overreport scales); and C&P had the highest rates for RDF, CBS, and SOS2 (again, all overreport scales). Within the overreporting scales, there was a 22% chance to elevate any given overreporting scale. Based on the mean percent of elevation for each SVT per stop code, the MFI had the highest mean at 30.8%, while the PDS had the lowest at 0.0%. For lowest elevations, the mental health biomedical clinics by far had the lowest number of elevations, with 14 across SVTs; in contrast, the

Table 3
Descriptive and Cumulative Base Rates for Supplementary Symptom Validity Scales

Scale/≥cutoff	Citation	*M*	*SD*	Min.–Max.	Kurtosis	Skew	Cum.%
MAL (n = 36,830)	Morey (2007)	1.15	1.11	0–8	1.03	0.97	
2/71 T							33.5
3/84 T							11.0
4/98 T							3.3
5/110 T							0.9
RDF (n = 36,823)	Morey (2007)	−0.66	1.24	−6.21–4.50	0.28	0.11	
0 raw/59 T							28.6
1.0 raw/69 T							8.9
DEF (n = 36,830)	Morey (2007)	1.66	1.58	0–9	0.01	0.81	
2/44 T							46.1
3/51 T							29.1
4/57 T							14.0
5/63 T							5.6
6/70 T							1.9
CDF (n = 36,814)	Morey (2007)	144.12	17.06	54.20–231.67	0.47	−0.17	
145							49.0
168							7.5
MFI (n = 36,830)	Gaines et al. (2013)	67.41	11.81	33.86–104.86	−0.49	−0.07	
74.01							30.4
76.01							24.8
CBS (n = 36,622)	Gaasedelen et al. (2019)	13.86	5.57	1–30	−0.64	0.07	
14							52.2
16							39.4
18							27.5
19							22.0
20							17.2
CBS SOS-1 (n = 36,818)	Boress et al. (2022)	68.67	13.44	37.83–117.83	−0.36	0.24	
78							25.0
CBS SOS-2 (n = 36,830)	Boress et al. (2022)	4.8	0.92	2.53–7.94	−0.45	0.13	
5.3							29.9
CBS SOS-3 (n = 36,830)	Boress et al. (2022)	67.56	12.01	37.29–111.29	−0.27	0.19	
74							29.6
HongM (n = 36,828)	Hong and Kim (2001)	−0.25	1.3	−3.46–4.96	−0.28	0.31	
1.59							9.3
HongD (n = 36,828)	Hong and Kim (2001)	−3.43	1.82	−9.19–2.68	−0.39	0.01	
0.92							0.4
HongR (n = 36,814)	Hong and Kim (2001)	−1.76	1.65	−4.63–4.90	0.19	0.87	
2.08							3.1
NDS (n = 36,507)	Mogge et al. (2010)	9.37	5.41	0–45	1.65	1.11	
11							34.7
13							24.1
15							16.3
19							6.5
25							1.6
27							0.9
PDS (n = 36,457)	Mogge et al. (2017)	18.63	5.88	0–51	0.67	0.67	
Raw 45							<.1

Note. For scales with more than one cutoff score, Grey color rows are cutoff score used in multivariate calculations. Min. = minimum; Max. = maximum; Cum. = cumulative; MAL = Malingering Index; RDF = Roger's Discriminant Function; DEF = Defensiveness Index; CDF = Cashel's Discriminant Function; MFI = Multiscale Feigning Index; CBS = Cognitive Bias Scale; SOS = Scale of Scales; HongM = Hong Malingering; HongD = Hong Defensiveness; HongR = Hong Randomness; NDS = Negative Distortion Scale; PDS = Positive Distortion Scale.

mental health individual clinic has 0 scales as the lowest percent. As far as scales functioning across clinics, the PDS had the lowest range of percentages at 0.4%, due to nearly all clinics showing 0.0% invalidity on the PDS; the HongD was a close second at 1.3%. In contrast, SOS1 had the highest range at 33.2%, with SOS3 falling second with 29.8%. Comparing invalid rates on any one or more SVT of the full 18 examined and three general contexts (forensic, clinical/testing, presurgical), disability clinics showed the highest rate of invalidity (66.3%), followed by the clinical testing clinic (54.1%), and lastly the presurgical (45.8%), supporting H6.

Discussion

All PAIs administered within the VA health care system and stored within CDW (N = 36,830) were obtained for the present

Table 4
Cumulative Frequencies by Various Scale Groupings

Group	Type	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14+
All[a]	All	38.4%	61.6%	42.3%	34.1%	28.6%	23.5%	18.7%	13.2%	8.6%	5.3%	2.5%	0.6%	<0.1%	<0.1%	<0.1%
	OR	43.5%	56.5%	38.6%	32.6%	27.8%	23.0%	18.0%	12.6%	8.1%	4.7%	1.9%				
	UR	96.3%	3.7%	0.7%	0.1%	0.0%	0.0%									
	NC	90.1%	9.9%	1.0%	0.1%											
Core 4[b]	All	81.9%	18.1%	2.0%	0.2%	0.0%										
	OR	89.9%	10.1%													
	UR	97.8%	2.2%													
	NC	92.8%	7.2%	0.6%												
Core 4+[c]	All	57.2%	42.8%	14.0%	4.2%	0.7%	<0.01%	0.0%	0.0%	0.0%						
	OR	62.6%	37.4%	9.6%	2.7%											
	UR	96.4%	3.6%	0.5%	0.0%											
	NC	92.8%	7.2%	0.6%												
Strong[d]	All	52.5%	47.5%	34.3%	26.5%	19.6%	10.6%	4.5%	0.5%	<0.01%	0.0%	0.0%				
	OR	59.5%	40.5%	32.4%	25.8%	19.0%	10.0%	3.9%								
	UR	96.4%	3.6%	0.5%												
	NC	92.8%	7.2%	0.6%												

Note. OR = overreport scales; UR = underreport scales; NC = noncontent-based scales; NIM = Negative Impression Management; PIM = Positive Impression Management; ICN = Inconsistency; INF = Infrequency; MAL = Malingering Index; RDF = Roger's Discriminant Function; DEF = Defensiveness Index; CDF = Cashel's Discriminant Function; CBS = Cognitive Bias Scale; SOS = Scale of Scales; PAI = Personality Assessment Inventory. [a] All group includes all 18 scales published to date. [b] Core 4 group includes main four scales profiled in standard profile: NIM, PIM, ICN, INF. [c] Core 4 + group includes Core 4 and supplementary scales in PAI Plus manual: NIM, MAL, RDF, PIM, DEF, CDF, ICN, INF. [d] Stron group includes scales with the strongest empirical support: NIM, MAL, CBS, SOS1, SOS2, SOS3, PIM, DEF, ICN, INF.

study. Thus, the current data set is the largest and most inclusive VA-related PAI data set available to date. In the present study, we conducted a descriptive analysis of multiple core and supplemental PAI validity scales to identify typical veteran response patterns and to further the knowledge of PAI validity scale outcomes within the VA. Our hypotheses and corresponding conclusions can be categorized into three general areas: (a) SVT base rates, (b) correlations among and discrepancies between select SVT scales, and (c) SVT failures as a function of test setting.

SVT Base Rates

There are a large number of PAI SVTs, and these scales differ substantially in empirical support and design. Below, we summarize

Table 5
Validity Scale Spearman Correlations

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. NIM	—																	
2. MAL	.68	—																
3. RDF	.22	.21	—															
4. MFI	.77	.63	.25	—														
5. CBS	.70	.54	.29	.84	—													
6. SOS1	.85	.65	.30	.93	.86	—												
7. SOS2	.80	.59	.22	.91	.88	.91	—											
8. SOS3	.83	.63	.30	.94	.88	.95	.96	—										
9. HM	.88	.65	.29	.88	.73	.86	.81	.86	—									
10. NDS	.72	.59	.19	.73	.65	.78	.73	.75	.72	—								
11. PIM	−.53	−.48	−.08	−.73	−.48	−.66	−.60	−.62	−.59	−.52	—							
12. DEF	−.43	−.27	−.10	−.60	−.44	−.59	−.53	−.54	−.49	−.38	.73	—						
13. CDF	.20	.19	.17	.24	.21	.16	.17	.21	.26	.20	.05	.15	—					
14. HD	−.59	−.53	−.43	−.79	−.67	−.79	−.70	−.77	−.64	−.55	.67	.60	−.08	—				
15. PDS	−.24	−.14	−.27	−.46	−.43	−.46	−.44	−.47	−.31	−.18	.31	.39	−.03	.50	—			
16. ICN	.08	−.06	.20	.03	.06	.07	.04	.06	.21	.03	.03	−.05	.01	−.03	−.08	—		
17. INF	.10	.04	.32	.04	.08	.07	.04	.07	.10	.10	.12	.07	.13	.02	−.15	.15	—	
18. HR	.31	.15	.15	.25	.16	.29	.21	.25	.37	.31	−.24	−.26	−.22	−.20	−.18	.44	.32	—

Note. Grey shading on upper left delineates overreport scales; Grey secondary shading underreport scales; no shading noncontent scales. NIM = Negative Impression Management; MAL = Malingering Index; RDF = Roger's Discriminant Function; MFI = Multiscale Feigning Index; CBS = Cognitive Bias Scale; SOS = Scale of Scales; HM = Hong Malingering; NDS = Negative Distortion Scale; PIM = Positive Impression Management; DEF = Defensiveness Index; CDF = Cashel's Discriminant Function; HD = Hong Defensiveness; PDS = Positive Distortion Scale; ICN = inconsistency; INF = infrequency; HR = Hong Randomness. See the online colour version of this table.

Table 6
Validity Scales by Clinic STOP Codes: Overreport Scales

Number	Name	*N*	NIM ≥92	MAL ≥3	RDF ≥0.1	MFI ≥75	CBS ≥19	SOS1 ≥78	SOS2 ≥5.3	SOS3 ≥74	HongM ≥1.59	NDS ≥13
197	Polytrauma	450	12.0%	10.4%	23.8%	28.7%	24.4%	22.0%	32.9%	28.9%	9.3%	32.1%
301	General internal medicine	103	15.5%	18.5%	30.1%	31.1%	27.2%	23.3%	30.1%	30.1%	16.5%	19.4%
502	MHC individual	5,603	10.3%	10.9%	26.7%	28.1%	20.4%	25.8%	28.1%	27.6%	9.3%	25.0%
509	Psychiatry MD individual	111	20.7%	11.7%	27.0%	37.8%	25.2%	38.7%	34.2%	33.3%	14.4%	34.6%
510	Psychology individual	1,203	11.6%	12.8%	25.2%	33.1%	21.6%	28.8%	31.8%	33.3%	11.2%	27.5%
512	Psychiatry consultation	128	8.6%	3.9%	22.7%	22.7%	20.3%	18.0%	22.7%	21.9%	5.5%	19.2%
513	Substance abuse individual	499	17.2%	13.4%	22.7%	36.5%	20.2%	31.7%	32.9%	33.1%	13.8%	33.3%
527	General psychiatry/phone	250	12.0%	14.8%	26.8%	34.0%	22.0%	29.6%	32.0%	33.2%	12.4%	27.9%
533	Mental health biomedical	201	1.0%	2.0%	21.4%	8.0%	9.5%	5.5%	11.0%	7.5%	1.5%	9.5%
534	Mental health integrated care	532	8.7%	8.5%	23.9%	21.2%	14.5%	16.5%	21.2%	19.0%	6.0%	17.7%
538	Psychological testing	9,581	6.9%	7.9%	24.6%	20.7%	15.1%	16.4%	21.9%	20.4%	5.9%	17.0%
540	PCT individual	821	11.3%	10.2%	30.5%	38.9%	26.4%	27.3%	35.0%	37.2%	10.5%	26.6%
550	MHC group	485	12.2%	11.1%	23.7%	28.0%	17.9%	27.0%	26.6%	26.4%	12.0%	29.7%
560	Substance abuse group	292	13.7%	14.0%	23.6%	34.3%	15.1%	28.1%	33.6%	32.5%	12.3%	34.3%
562	PTSD individual	1,446	12.4%	12.2%	27.9%	39.8%	26.8%	30.0%	35.7%	37.3%	12.2%	28.0%
582	PRRC individual	108	17.6%	16.7%	31.5%	30.6%	25.9%	33.3%	30.6%	33.3%	11.1%	32.7%
586	RRTP individual	912	16.6%	14.3%	23.5%	37.1%	18.0%	30.7%	31.4%	33.0%	15.0%	32.7%
587	RRTP group	534	17.6%	14.2%	24.2%	35.8%	22.7%	30.7%	33.3%	34.1%	16.3%	33.2%
[a]	Disability exam	10,752	10.8%	13.1%	35.1%	37.9%	28.6%	30.3%	37.0%	37.2%	10.5%	26.7%
	Mean %		12.5%	11.6%	26.0%	30.8%	21.1%	26.0%	29.6%	29.4%	10.8%	26.7%
	Lowest %		1.0%	2.0%	21.4%	8.0%	9.5%	5.5%	11.0%	7.5%	1.5%	9.5%
	Highest %		20.7%	18.5%	35.1%	39.8%	28.6%	38.7%	37.0%	37.3%	16.5%	34.6%
	Range		19.7%	16.5%	13.7%	31.8%	19.1%	33.2%	26.0%	29.8%	15.0%	25.1%

Note. Grey cells indicate the highest rate per validity measure (per column; e.g., highest invalid rate on NIM was in stop code 509). NIM = Negative Impression Management; MAL = Malingering Index; RDF = Roger's Discriminant Function; MFI = Multiscale Feigning Index; CBS = Cognitive Bias Scale; SOS = Scale of Scales; HongM = Hong Malingering; NDS = Negative Distortion Scale; MD = doctor of medicine; MHC = Mental Health Clinic; PCT = PTSD Clinical Team; PTSD = posttraumatic stress disorder; PRRC = Psychosocial Rehabilitation and Recovery Center; RRTP = Residential Rehabilitation Treatment Program.
[a] Disability exam status included all protocols administered in clinics with secondary stop codes including 179 (Telehealth), 443 (Disability Benefits Questionnaire referral face to face), 444/445/445/447 (telehealth), 448 (Integrated Disability Evaluation System exam), 450 (Compensation & Pension), and 697 (Acceptable Clinical Evidence Compensation & Pension).

several patterns of invalidity observed across PAI SVTs, doing so based on different groupings of scales. We start by focusing on the core overreport (NIM), underreport (PIM), and noncontent-based (ICN and INF) scales. Across these data, 18.1% of administrations were invalid based on these criteria, with the highest frequency of invalidity by overreporting (10.1%), followed by noncontent-based invalidity (7.2%), and then underreporting (2.2%) (H1 and H2). This trend continued when the major supplemental scales were added to the overreport and underreport categories (i.e., MAL/RDF and DEF/CDF, respectively), with 37.4% for overreporting, 7.2% for underreporting, and 3.6% for noncontent-based responding. When considering all primary and supplemental scales, roughly 62% of administrations demonstrated elevated scores on one or more overreporting, underreporting, or noncontent-based invalidity scale, and 42% for 2 + SVTs. When evaluated by all PAI SVTs, 56.5% of profiles were invalided by overreport, 9.9% by noncontent-based responding, and 3.7% by underreport. Because disability evaluations have the highest invalidity rates and represent the largest proportion of the overall sample, those rates influence the overall infrequency rate. Interestingly, this is the same pattern of results found when analyzing MMPI-2-RF validity scores from the VA population (H3). Per Ingram et al. (2020b), approximately 37.2% of MMPI-2-RF protocols were invalid due to overreporting (12.6% for single SVT elevations and 24.6% with multiple elevations), 4.9% were invalid due to noncontent-based invalidity, and 2.5% were invalid due to underreporting with an overall rate of invalidity of 57.8%. Treating each stop code as an independent sample allows to control effects of unequal sample sizes. The average elevation of elevation on a single overreporting scales is about 22%, which is slightly higher than the approximate 13.9% seen by Ingram et al. (2020b) on the MMPI-2-RF. Taken together, these trends provide further evidence that overreporting is the most common form of invalidity within VA settings (H2) and that underreporting and noncontent responding are, in comparison, relatively rare.

Rates of overreporting were variable, depending on the scale considered. When invalidity was determined by the NIM, MAL, and HongM, rates of invalidity were around 10%. Conversely, rates were in the 25% to 30% range when measured with RDF, MFI, NDS, CBS, SOS-1, SOS-2, and SOS-3. Ingram et al. (2020b) found that although scales differ fairly substantively on elevation rate at recommend cut scores (symptom validity = 5.2% to Response Bias Scale = 27.3%), base rates of failed overreporting scales in the VA are approximately 37.2% when measured by failure on any MMPI-2-RF validity scale. Invalid cognitive performance in VA health care settings is around 30% (Shura et al., 2022), consistent with Young et al.'s (2025) suggestion that a tentative rate of 30% invalid is a reasonable expectation. Accordingly, our findings suggest that the standard cutoffs for NIM, MAL, and HongM might be too conservative in VA health care settings if matching for base rates. Ongoing research examining alternative cutoffs within the VA health care setting is potentially warranted, particularly for NIM, MAL, and HongM. It is also noteworthy that the cognitive overreport scales in

Table 7
Validity Scales by Clinic STOP Codes: Underreport and Noncontent Scales

Number	Name	*N*	PIM ≥68	DEF ≥6	CDF ≥168	HongD ≥0.92	PDS ≥45	ICN ≥73	INF ≥75	HongR ≥2.08	Any SVT ≥1
197	Polytrauma	450	3.1%	1.8%	9.3%	0.0%	0.0%	4.2%	4.0%	0.4%	59.6%
301	General internal medicine	103	4.9%	4.9%	8.7%	1.0%	0.0%	3.9%	2.9%	0.0%	58.3%
502	MHC individual	5,603	2.8%	2.7%	9.1%	0.6%	0.1%	4.8%	3.3%	2.1%	61.1%
509	Psychiatry MD individual	111	0.0%	0.0%	3.6%	0.0%	0.0%	6.3%	7.2%	5.4%	68.5%
510	Psychology individual	1,203	3.6%	2.2%	10.0%	0.8%	0.0%	4.4%	3.5%	6.2%	64.3%
512	Psychiatry consultation	128	1.6%	1.6%	8.6%	0.0%	0.0%	7.0%	4.7%	0.8%	48.4%
513	Substance abuse individual	499	1.2%	0.8%	5.2%	1.0%	0.0%	3.8%	3.6%	17.4%	66.7%
527	General psychiatry/phone	250	2.4%	1.2%	11.6%	0.8%	0.0%	6.0%	4.0%	1.6%	61.6%
533	Mental health biomedical	201	7.5%	3.5%	3.0%	1.0%	0.0%	3.0%	0.5%	0.0%	45.8%
534	Mental health integrated care	532	3.4%	4.0%	7.0%	1.3%	0.0%	2.8%	3.4%	0.6%	54.1%
538	Psychological testing	9,581	2.7%	2.1%	7.8%	0.5%	0.0%	5.0%	4.0%	1.6%	54.1%
540	PCT individual	821	0.6%	1.0%	10.4%	0.4%	0.0%	4.3%	3.9%	2.1%	67.0%
550	MHC group	485	2.7%	2.3%	7.4%	1.0%	0.0%	7.6%	6.6%	7.9%	64.3%
560	Substance abuse group	292	0.7%	2.4%	3.8%	1.0%	0.4%	3.4%	5.1%	28.4%	72.6%
562	PTSD individual	1,446	0.9%	0.6%	10.9%	0.0%	0.0%	4.4%	2.8%	1.9%	64.4%
582	PRRC individual	108	2.8%	0.0%	5.6%	0.9%	0.0%	4.6%	7.4%	2.8%	69.4%
586	RRTP individual	912	0.7%	0.3%	4.3%	0.3%	0.0%	3.3%	3.8%	19.9%	64.9%
587	RRTP group	534	0.9%	0.0%	3.6%	0.2%	0.0%	4.1%	4.1%	18.4%	68.0%
[a]	Disability exam	10,752	1.0%	1.2%	9.2%	0.1%	0.0%	3.9%	2.7%	1.3%	66.30%
	Mean %		2.3%	1.7%	7.3%	0.6%	0.0%	4.6%	4.1%	6.3%	62.1%
	Lowest %		0.0%	0.0%	3.0%	0.0%	0.0%	2.8%	0.5%	0.0%	45.8%
	Highest %		7.5%	4.9%	11.6%	1.3%	0.4%	7.6%	7.4%	28.4%	72.6%
	Range		7.5%	4.9%	8.6%	1.3%	0.4%	4.8%	6.9%	28.4%	26.8%

Note. Grey cells indicate the highest rate per validity measure (per column; e.g., highest invalid rate on PIM was in stop code 533). PIM = Positive Impression Management; DEF = Defensiveness Index; CDF = Cashel's Discriminant Function; HongD = Hong Defensiveness; PDS = Positive Distortion Scale; ICN = Inconsistency; INF = Infrequency; HongR = Hong Randomness; MD = doctor of medicine; MHC = Mental Health Clinic; PCT = PTSD Clinical Team; PTSD = posttraumatic stress disorder; PRRC = Psychosocial Rehabilitation and Recovery Center; RRTP = Residential Rehabilitation Treatment Program; SVT = symptom validity test.
[a] Disability exam status included all protocols administered in clinics with secondary stopcodes including 179 (Telehealth), 443 (Disability Benefits Questionnaire referral face to face), 444/445/445/447 (telehealth), 448 (Integrated Disability Evaluation System exam), 450 (Compensation & Pension), and 697 (Acceptable Clinical Evidence Compensation & Pension). The Any SVT column included all SVTs (including overreport scales from Table 6).

general were higher than most other overreport scales (with RDF being an exception), which could suggest that overreporting cognitive complaints is more common in this population than overreporting psychopathology. This finding is again consistent with Ingram and colleagues (Ingram et al., 2020b); where the cognitive-based Response Bias Scale was the most elevated overreport scale at 27.3% (followed by F-r at 23.2%, a measure of psychopathology overreporting). This finding compares well to the range of invalid cognitive complaints on the four cognitively based overreport SVTs (e.g., CBS/SOSs [22.0% to 29.9%]). Our results highlight the importance of sampling overreporting in veterans and the need to consider overreporting of cognitive complaints in each exam.

Rates of underreporting were below 5% for all pertinent validity indices except for CDF (7.5%). When considering that base rates of underreporting were 2.5% when measured with the MMPI-2-RF validity scales (Ingram et al., 2020b), this suggests that the CDF cutoff might require modification within VA health care settings. Alternatively, as outlined below, correlational analyses also suggest that the CDF may not operate particularly well in this population. More broadly, and potentially problematically, underreporting scales often have low sensitivity. They also rarely elevate in a manner that would identify under reporters within simulation studies with veterans (Khazem et al., 2025) and other populations (Crighton et al., 2017).

Finally, rates of noncontent-based invalidity were 4.5% for ICN and 3.4% for INF, totaling 7.2% when combined. HongR demonstrated a similar rate of noncontent-based invalidity, with 3.1% of individuals invalidating this scale. Considering all three scales, 9.9% of the sample engaged in noncontent responding across the full sample. Full sample invalidity for these scales (INF, ICN, HongR) is twice that identified from the MMPI-2-RF (4.9% overall) in a large VA sample (Ingram et al., 2020b). These rates were more elevated in general during analyses of different service clinics, particularly for INF (4.1% mean elevation) and HongR (6.3% mean elevation). It is unclear what is driving the difference between PAI and MMPI-2-RF invalidity rates for noncontent-related scales, but one possibility relates to different scale construction methods. For example, ICN is similar to variable response inconsistency and true response inconsistency in that it looks at consistency across items; however, INF is closer to F-r in that it is comprised of infrequently endorsed items, half of which are expected to be answer false, not at all t, the other half very true. Items are rare, but not necessarily contradictory. INF is therefore more sensitive to extreme responding and may reflect broader positive and negative response bias detection, more like a content-based scale (e.g., F-r) than a noncontent counterpart (variable response inconsistency and true response inconsistency). Similarly, HongR uses discriminant function (i.e., β weights) designed to predict random responders; of the four scales it uses, two of them are in fact ICN and INF (the other two alcohol problems and drug problems). These differences in scale construction might be extending these results to a construct other than random noncontent-based response, in contrast to the MMPI-2-RF results.

Correlations Among and Discrepancies Between Select SVTs

When examining the overreport validity scales, correlations were strong, with the sole exception being the RDF (H5). Outside of RDF, correlations ranges from .54 (CBS and MAL) to .96 (SOS2 and SOS3). RDF correlations, however, ranges from .19 (NDS) to .30 (SOS1 and SOS3). This suggests that RDF, which was originally validated with simulation studies, might not be a useful overreport validity scale in clinical settings, a conclusion also offered by others (Boccaccini & Hart, 2018).

The strong positive correlations among PAI overreport scales (save RDF) could be taken to suggest that these scales might assess a common underlying construct, one that constitutes a general tendency to overreport (H5). This includes strong correlations between scales purporting to measure psychological overreporting and cognitive overreporting. For instance, NIM (arguably the most robust PAI measure of overreported psychological symptoms) and the CBS (arguably the most robust PAI measure of overreported cognitive symptoms) correlate strongly at .70 suggesting somewhat limited unique variance. A correlation of .70 indicates 49% shared variance, leaving 51% to represent both systematic (e.g., method variance, evaluation setting variance, distress-driven endorsement, etc.) and random measurement error, as well as cognitive, somatic, and psychological-specific variance.

Potentially further supporting this hypothesis, the overreporting validity scales have uniformly negative correlations with most underreporting validity scales and weak correlations with most noncontent-based validity scales. Our assertion that there may be a single underlying construct assessed by overreport validity scales does not appear to be specific to our sample or our VA setting. Indeed, some researchers have suggested a more general overreporting style serves as the primary elevating factor for SVTs (Ingram et al., 2024; Schroeder et al., 2025). For instance, correlations between SVTs with purportedly different domains are fairly consistent with the values ($r \sim .70$) observed here when measured in civil and criminal (Wygant et al., 2010) or veteran (Schroeder et al., 2025) samples. Research using multivariate base rates (e.g., Aita et al., 2023) has also recognized this underlying principle across invalidated protocols (e.g., elevation during overreporting is broad and occurs in volume, generally without respect to domain or scale content). Likewise, the creation of the MFI stemmed from the shared observation of the researchers that scales elevated broadly and without respect to symptom expectation (Gaines et al., 2013).

Still, it is acknowledged that cognitive overreport scales elevated at higher rates in both the current data and previous MMPI-2-RF data (Ingram et al., 2020b). To this end, it is possible that there is a primary factor related to overreporting and then secondary factors related to overreport domain. Given the inclusion of distinct domains in SVT interpretation standards (Sweet et al., 2021), research examining the distinctions between overreporting domains (e.g., cognitive vs. psychological) should be pursued with additional methods such as structural equation modeling and latent class techniques.

Regarding the underreport validity scales, a similar trend emerged as seen with the overreporting scales. Specifically, most correlations between the underreporting scales were strong, but PDS correlations with other underreporting scales were (at best) moderate, and CDF correlations were weak. In fact, CDF is the only underreporting validity scale that did not consistently positively correlate with the other underreporting validity scales, suggesting that CDF may be of limited value in assessing underreporting within veterans.

Finally, there appears to be some degree of variability in the noncontent-based validity scales. While these scales frequently demonstrate weak correlations with both the overreporting and underreporting validity scales (suggesting that they are likely measuring a different type of validity), the strength of the noncontent-based validity scales ranges from moderate (0.44; ICN and HongR) to weak (0.15; ICN and INF) when examined together. Such an observation suggests that different underlying constructs could potentially be measured by the noncontent-based validity scales within a VA setting. However, further evidence to this end is clearly necessary to confidently draw such a conclusion.

SVT Failures as a Function of Test Setting

In describing these data, three test settings seem to emerge as particularly relevant: (a) disability evaluations, (b) presurgical evaluations, and (c) general clinical practice. Within the VA, the PAI was administered frequently in disability exam settings ($n = 10{,}752$). In evaluating the core overreporting scales, 10.8% elevated the NIM (≥92) and 13.1% elevated the MAL (≥3). Of note, a substantially higher number elevated the CBS and CB-SoSs (ranging from 28.6% to 37.2%), fairly consistent with rates observed for Response Bias Scale (Ingram et al., 2022). Conversely, in this setting, only 1% invalided that PAI secondary to underreporting (per the PIM; H6). Higher rates of overreport within disability examinations are predictable given the clear external incentives present in these evaluations. Indeed, Shura and colleagues (Shura et al., 2022) reported that VA disability evaluations are associated with higher rates of performance invalidity relative to clinical settings. Comparable patterns are seen on the MMPI-2-RF, with compensation-focused evaluations producing more invalidity, at a small to moderate magnitude of effect (Diehl et al., 2025). Thus, the high rates of invalidity may be related to concerted efforts to appear more pathological than clinically justified to achieve a higher disability rating.

On the other end of the continuum, the setting that has the lowest failure percentage on the most overreporting validity scales was that of presurgical evaluations (presurgical transplant evaluations, presurgical bariatric evaluations, etc.), represented by the Mental Health Biomedical stop code. This clinic also had the highest rate of underreporting as gauged by the PIM. Such a configuration of very low overreporting and higher underreporting makes conceptual sense, given that individuals who undergo presurgical evaluations have an incentive to look well-adjusted as to avoid any barriers to engaging the proposed surgery (H6).

Regarding routine clinical mental health care/evaluation, the Psychological Testing stop code is likely the best representation, and this stop code was associated with nearly as many administrations of PAIs ($n = 9{,}851$) as the disability examinations. Somewhat predictably, rates of invalidity based on overreport and underreport fell between the disability and presurgical settings (H6). The finding that overreporting scales are most frequently elevated in disability evaluations is consistent with previous literature (Diehl et al., 2025; Ingram et al., 2020b; Shura et al., 2022) and expectation, given (in part) the forensically enmeshed nature of those evaluations (Nelson et al., 2010; Ray, 2017; Worthen & Moering, 2011).

Of additional note, there were high rates of overreport invalidity in the Psychiatric medical doctor Individual clinic setting. The fact that medical doctors are administering PAIs and given that only 111 PAIs (i.e., <0.5% of all administered PAIs within the VA) have been administered in this setting to date suggests that the individuals who are being administered the PAIs in this setting are a rare and select group of examinees. We speculate that these 111 individuals were targeted by the medical doctors because something about the case was idiosyncratic or suspicious. Thus, there may have been a selection bias whereby only individuals who were so atypically presenting at baseline were administered testing. Many of the other clinics outside the three mentioned appear to reflect unique or nuanced settings, and as such, invalidity rates were quite variable. These clinics reflect a variety of conditions (serious mental illness, PTSD, substance use, traumatic brain injury), providers, settings, and administration conditions (individual vs. group). These remain included in the event readers are working with similar subpopulations for comparison purposes.

In summary, when looking at clinic setting data, overreporting is likely to occur in (a) settings where there are clear external incentives or (b) settings where atypical presentations are likely. Overreporting is somewhat less likely to occur in general clinical settings, although this style of response bias still occurs with some frequency. Conversely, overreporting is uncommon in settings where there are external incentives to look nonpathological (e.g., presurgical settings). This is intuitive given prior research findings, but it is now clearly visible within VA health care settings based on the current data (H6).

Strengths and Limitations

The trends described throughout this article are believed to be the most accurate representation of typical VA-based PAI performances to date because they incorporate every PAI administered within the VA health care system and stored within CDW. Thus, the data can be considered a valid representation of validity scale elevations within the VA health care system. While there are clear strengths to this study, there are also limitations. One limitation is that we did not examine external criteria to assess classification accuracy values of the PAI validity scales. Consequently, we are unable to make certain conclusions regarding reasons behind elevated validity scale scores. We do speculate on this, though, and encourage future research to follow-up and examine whether our hypotheses are accurate. Some statistical and methodological questions have been raised about how these same statistics are interpreted and leveraged to make decisions about interpretation (Leonhard, 2023a, 2023b; Niesten et al., 2022), underscoring the need for expanded study. Future research would also benefit from evaluating the degree to which any diagnosis might affect specific validity scale scores. Additionally, findings may not generalize to other nonveteran populations, and additional research on diverse populations would benefit this line of study.

References

Aita, S. L., Moncrief, G. G., Greene, J., Trujillo, S., Carrillo, A., Iwanicki, S., Morera, C. C., Gioia, G. A., Isquith, P. K., & Roth, R. M. (2023). Univariate and multivariate base rates of score elevations, reliable change, and interrater discrepancies in the BRIEF-A standardization samples. *Assessment*, *30*(2), 390–401. https://doi.org/10.1177/10731911211055673

Apaydin, E. A., Paige, N. M., Begashaw, M. M., Larkin, J., Miake-Lye, I. M., & Shekelle, P. G. (2023). Veterans Health Administration (VA) vs. Non-VA healthcare quality: A systematic review. *Journal of General Internal Medicine*, *38*(9), 2179–2188. https://doi.org/10.1007/s11606-023-08207-2

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of Psychological Test use patterns among forensic psychologists. *Journal of Personality Assessment*, *87*(1), 84–94. https://doi.org/10.1207/s15327752jpa8701_07

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

Asch, S. M., McGlynn, E. A., Hogan, M. M., Hayward, R. A., Shekelle, P., Rubenstein, L., Keesey, J., Adams, J., & Kerr, E. A. (2004). Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample. *Annals of Internal Medicine*, *141*(12), 938–945. https://doi.org/10.7326/0003-4819-141-12-200412210-00010

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

Blais, M. A., Baity, M. R., & Hopwood, C. J. (2010). *Clinical applications of the Personality Assessment Inventory*. Routledge/Taylor & Francis Group; https://research.ebsco.com/linkprocessor/plink?id=5d8a506b-e9a0-3bf5-b2fa-6804e372c768

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

Boccaccini, M. T., & Hart, J. R. (2018). Response style on the Personality Assessment Inventory and other multiscale inventories. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 280–300). Guilford Press. https://research.ebsco.com/linkprocessor/plink?id=d893e7a5-e6c2-3bcc-b2db-c9be4acaff9d

Boress, K., Gaasedelen, O. J., Croghan, A., Johnson, M. K., Caraher, K., Basso, M. R., & Whiteside, D. M. (2022). Validation of the Personality Assessment Inventory (PAI) scale of scales in a mixed clinical sample. *The Clinical Neuropsychologist*, *36*(7), 1844–1859. https://doi.org/10.1080/13854046.2021.1900400

Braxton, L. E., Calhoun, P. S., Williams, J. E., & Boggs, C. D. (2007). Validity rates of the Personality Assessment Inventory and the Minnesota Multiphasic Personality Inventory-2 in a VA Medical Center setting. *Journal of Personality Assessment*, *88*(1), 5–15. https://doi.org/10.1080/00223890709336829

Breshears, R. E., Brenner, L. A., Harwood, J. E. F., & Gutierrez, P. M. (2010). Predicting suicidal behavior in veterans with traumatic brain injury: The utility of the Personality Assessment Inventory. *Journal of Personality Assessment*, *92*(4), 349–355. https://doi.org/10.1080/00223891.2010.482011

Burchett, D., & Ben-Porath, Y. S. (2019). Methodological considerations for developing and evaluating response bias indicators. *Psychological Assessment*, *31*(12), 1497–1511. https://doi.org/10.1037/pas0000680

Calhoun, P. S., Collie, C. F., Clancy, C. P., Braxton, L. E., & Beckham, J. C. (2010). Use of the PAI in assessment of posttraumatic stress disorder among help-seeking veterans. In M. A. Blais, & C. J. Hopwood (Eds.), *Clinical applications of the Personality Assessment Inventory* (pp. 93–111). Routledge/Taylor & Francis Group. https://research.ebsco.com/linkprocessor/plink?id=2115fabc-e691-3a1b-b73a-211bc3fda811

Cersosimo, B. H., Hilsenroth, M. J., Bornstein, R. F., Gold, J. R., & Blais, M. A. (2022). Personality Assessment Inventory items in relation to patient- and therapist-rated alliance. *Clinical Psychology & Psychotherapy*, *29*(6), 1905–1917. https://doi.org/10.1002/cpp.2759

Crawford, E. F., Calhoun, P. S., Braxton, L. E., & Beckham, J. C. (2007). Validity of the Personality Assessment Inventory aggression scales and

Violence Potential Index in veterans with PTSD. *Journal of Personality Assessment*, *88*(1), 90–98. https://doi.org/10.1080/00223890709336839

Crighton, A. H., Marek, R. J., Dragon, W. R., & Ben-Porath, Y. S. (2017). Utility of the MMPI-2-RF validity scales in detection of simulated underreporting: Implications of incorporating a manipulation check. *Assessment*, *24*(7), 853–864. https://doi.org/10.1177/1073191115627011

Denning, J. H., & Horner, M. D. (2024). The impact of race and other demographic factors on the false positive rates of five embedded Performance Validity Tests (PVTs) in a Veteran sample. *Journal of Clinical and Experimental Neuropsychology*, *46*(1), 25–35. https://doi.org/10.1080/13803395.2024.2314737

Department of Veteran Affairs. (2024). *National Center for Veterans Analysis and Statistics: Veteran population*. https://www.va.gov/vetdata/veteran_population.asp

Diehl, K. J., Ingram, P. B., Pagano, L. A., & Gideon, H. J. (2025). *Patterns of MMPI-2-RF validity scale elevation across veterans seen in a VA assessment clinic: The impact of compensation status*. Psychological Services.

Fihn, S. D., Francis, J., Clancy, C., Nielson, C., Nelson, K., Rumsfeld, J., Cullen, T., Bates, J., & Graham, G. L. (2014). Insights from advanced analytics at the Veterans Health Administration. *Health Affairs*, *33*(7), 1203–1211. https://doi.org/10.1377/hlthaff.2014.0054

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

Glenn, D. M., Beckham, J. C., Sampson, W. S., Feldman, M. E., Hertzberg, M. A., & Moore, S. D. (2002). MMPI-2 profiles of Gulf and Vietnam combat veterans with chronic posttraumatic stress disorder. *Journal of Clinical Psychology*, *58*(4), 371–381. https://doi.org/10.1002/jclp.1149

Haley, R. W., Kurt, T. L., & Hom, J. (1997). Is there a Gulf War Syndrome? Searching for syndromes by factor analysis of symptoms. *JAMA*, *277*(3), 215–222. https://doi.org/10.1001/jama.1997.03540270041025

Harrison, A. G., Harrison, K. A., & Armstrong, I. T. (2022). Discriminating malingered attention deficit hyperactivity disorder from genuine symptom reporting using novel Personality Assessment Inventory validity measures. *Applied Neuropsychology. Adult*, *29*(1), 10–22. https://doi.org/10.1080/23279095.2019.1702043

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. https://doi.org/10.1037/a0015036

Hong, S. W., & Kim, Y. H. (2001). Detection of random response and impression management in the PAI: II. Detection indices. *Korean Journal of Clinical Psychology*, *20*, 751–761.

Ingram, P. B., Keen, M. A., Greene, T. E., Morris, C., & Armistead-Jehle, P. J. (2024). Development and initial validation of the Scale of Scales (SOS) overreporting scores for the MMPI family of instruments. *Journal of Clinical and Experimental Neuropsychology*, *46*(2), 95–110. https://doi.org/10.1080/13803395.2024.2320453

Ingram, P. B., Morris, C. S., 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*(3), 624–635. https://doi.org/10.1007/s10880-021-09812-1

Ingram, P. B., Tarescavage, A. M., Ben-Porath, Y. S., & Oehlert, M. E. (2020a). Comparing MMPI-2-restructured form scores by service era for veterans assessed within the veteran affairs healthcare system. *Journal of Clinical Psychology in Medical Settings*, *27*(2), 366–375. https://doi.org/10.1007/s10880-019-09650-2

Ingram, P. B., Tarescavage, A. M., Ben-Porath, Y. S., & Oehlert, M. E. (2020b). Patterns of MMPI-2-Restructured Form (MMPI-2-RF) validity scale scores observed across Veteran Affairs settings. *Psychological Services*, *17*(3), 355–362. https://doi.org/10.1037/ser0000339

Kennedy, J. E., Cooper, D. B., Reid, M. W., Tate, D. F., & Lange, R. T. (2015). Profile analyses of the Personality Assessment Inventory following military-related traumatic brain injury. *Archives of Clinical Neuropsychology*, *30*(3), 236–247. https://doi.org/10.1093/arclin/acv014

Kessler, R. C., Heeringa, S. G., Stein, M. B., Colpe, L. J., Fullerton, C. S., Hwang, I., Naifeh, J. A., Nock, M. K., Petukhova, M., Sampson, N. A., Schoenbaum, M., Zaslavsky, A. M., Ursano, R. J., & the Army STARRS Collaborators. (2014). Thirty-day prevalence of DSM-IV mental disorders among nondeployed soldiers in the US Army: Results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*, *71*(5), 504–513. https://doi.org/10.1001/jamapsychiatry.2014.28

Khazem, L., Keen, M., Rodriguez, T. R., Ingram, P. B., Hay, J. M., Long, C. M., Bryan, C. J., & Anestis, J. C. (2025). Detecting simulated underreporting on the Minnesota Multiphasic Personality Inventory-3 (MMPI-3) in veterans with past-month death/suicide ideation. *Suicide and Life-Threatening Behavior*, *55*(2), Article e13170. https://doi.org/10.1111/sltb.13170

Kilpatrick, D. G., Resnick, H. S., Milanak, M. E., Miller, M. W., Keyes, K. M., & Friedman, M. J. (2013). National estimates of exposure to traumatic events and PTSD prevalence using DSM-IV and DSM-5 criteria. *Journal of Traumatic Stress*, *26*(5), 537–547. https://doi.org/10.1002/jts.21848

Kümmel, G. (2018). Military identity and identity within the military. In G. Caforio & M. Nuciari (Eds.), *Handbook of the sociology of the military* (pp. 477–493). Springer International Publishing. https://doi.org/10.1007/978-3-319-71602-2_25

Kurtz, J. E., & Blais, M. A. (2007). Introduction to the special issue on the Personality Assessment Inventory. *Journal of Personality Assessment*, *88*(1), 1–4. https://doi.org/10.1080/00223890709336828

Kurtz, J. E., & McCredie, M. N. (2022). Exaggeration or fabrication? Assessment of negative response distortion and malingering with the Personality Assessment Inventory. *Psychological Injury and Law*, *15*(1), 37–47. https://doi.org/10.1007/s12207-021-09433-x

Kurtz, J. E., & Pintarelli, E. M. (2024). The Daubert standards for admissibility of evidence based on the Personality Assessment Inventory. *Psychological Injury and Law*, *17*(2), 105–116. https://doi.org/10.1007/s12207-024-09508-5

Leonhard, C. (2023a). Review of statistical and methodological issues in the forensic prediction of malingering from validity tests: Part I: Statistical issues. *Neuropsychology Review*, *33*(3), 581–603. https://doi.org/10.1007/s11065-023-09601-7

Leonhard, C. (2023b). Review of statistical and methodological issues in the forensic prediction of malingering from validity tests: Part II—Methodological issues. *Neuropsychology Review*, *33*(3), 604–623. https://doi.org/10.1007/s11065-023-09602-6

MacLean, A., & Elder, G. H., Jr. (2007). Military service in the life course. *Annual Review of Sociology*, *33*(1), 175–196. https://doi.org/10.1146/annurev.soc.33.040406.131710

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2025). Neuropsychological validity assessment beliefs and practices: A survey of North American neuropsychologists and validity assessment experts. *Archives of Clinical Neuropsychology*, *40*(2), 201–223. https://doi.org/10.1093/arclin/acae102

McCredie, M. N., Kurtz, J. E., & Valentine, L. (2018). Prediction of psychotherapy process and outcome with the Personality Assessment

Inventory. *Psychiatry Research*, *269*, 455–461. https://doi.org/10.1016/j.psychres.2018.08.110

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

Meaux, L., Cox, J., Edens, J. F., DeMatteo, D., Martinez, A., & Bownes, E. (2022). The Personality Assessment Inventory in U.S. case law: A survey and examination of relevance to legal proceedings. *Journal of Personality Assessment*, *104*(2), 179–191. https://doi.org/10.1080/00223891.2021.1975723

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

Mogge, N. L., Hong, S. H., & Morey, L. C. (2017). *The Positive Distortion Scale (PDS): A study of a new PAI validity scale* [Unpublished Manuscript]. Department of Behavioral Medicine and Psychiatry, West Virginia University.

Mogge, N. L., Lepage, J. S., Bell, T., & Ragatz, L. (2010). The Negative Distortion Scale: A new PAI validity scale. *Journal of Forensic Psychiatry & Psychology*, *21*(1), 77–90. https://doi.org/10.1080/14789940903174253

Morey, L. C. (2007). *Personality Assessment Inventory: Professional manual*. Psychological Assessment Resources.

Morey, L. C. (2020). *PAI Plus: Professional manual supplement*. Psychological Assessment Resources.

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

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

Nelson, N. W., Anderson, C. R., Hoelzle, J. B., & Arbisi, P. A. (2014). Psychological assessment of veterans in outpatient mental health settings. In S. S. Bush (Ed.), *Psychological assessment of veterans* (pp. 17–50). Oxford University Press. https://research.ebsco.com/linkprocessor/plink?id=10997bdc-4050-390b-8665-45928ed45d9e

Nelson, N. W., Hoelzle, J. B., McGuire, K. A., Ferrier-Auerbach, A. G., Charlesworth, M. J., & Sponheim, S. R. (2010). Evaluation context impacts neuropsychological performance of OEF/OIF veterans with reported combat-related concussion. *Archives of Clinical Neuropsychology*, *25*(8), 713–723. https://doi.org/10.1093/arclin/acq075

Niesten, I. J. M., Merckelbach, H., Dandachi-FitzGerald, B., Jutten-Rooijakkers, I., & van Impelen, A. (2022). Experts' failure to consider the negative predictive power of symptom validity tests. *Frontiers in Psychology*, *13*, Article 789762. https://doi.org/10.3389/fpsyg.2022.789762

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

Rogers, R., & Bender, S. D. (2018). *Clinical assessment of malingering and deception*. Guilford Press. https://research.ebsco.com/linkprocessor/plink?id=48d75c95-8e42-3ab7-9316-1c597acad479

Ruchensky, J. R., Kelley, S. E., Massey, C., Richardson, L. A., Blais, M. A., & Stein, M. B. (2024). Using the Personality Assessment Inventory to assess the alternative model for personality disorders: Criterion validity in a clinical sample. *Journal of Personality Assessment*, *106*(1), 72–82. https://doi.org/10.1080/00223891.2023.2203240

Russell, D. N., & Morey, L. C. (2019). Use of validity indicators on the Personality Assessment Inventory to detect feigning of post-traumatic stress disorder. *Psychological Injury and Law*, *12*(3–4), 204–211. https://doi.org/10.1007/s12207-019-09349-7

Russo, A. C. (2014). Assessing veteran symptom validity. *Psychological Injury and Law*, *7*(2), 178–190. https://doi.org/10.1007/s12207-014-9190-2

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

Schroeder, R. W., Bieu, R. K., & Snodgrass, M. (2025). Comparing the cognitive bias scale and cognitive bias scale of scales to other Personality Assessment Inventory validity scales for detecting noncredible memory dysfunction in a clinical veteran sample. *Journal of Clinical and Experimental Neuropsychology*, *47*(1–2), 12–25. https://doi.org/10.1080/13803395.2025.2464635

Schroeder, R. W., Boone, K. B., & Larrabee, G. J. (2021). Design methods in neuropsychological performance validity, symptom validity, and malingering research. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (2nd ed., pp. 11–33). Guilford Press. https://research.ebsco.com/linkprocessor/plink?id=e66cef7c-0570-31ae-b98e-b6b038c25dcb

Shura, R. D., Denning, J. H., Armistead-Jehle, P., & Axelrod, B. N. (2022). Validity assessment within veterans affairs and active duty settings. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 235–253). Guilford Press. https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2021-84988-013&site=ehost-live

Shura, R. D., Ingram, P. B., Miskey, H. M., Martindale, S. L., Rowland, J. A., & Armistead-Jehle, P. (2023). 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*, *37*(7), 1548–1565. https://doi.org/10.1080/13854046.2022.2131630

Shura, R. D., Miskey, H. M., McDonald, S., Rowland, J. A., Martindale, S. L., Wagner, H. R., & Tupler, L. A. (2023). Personality Assessment Inventory suicidality scales: Suicidal Ideation (SUI), Suicide Potential Index (SPI), and S_Chron in an Afghanistan/Iraq-era active and Veteran military sample. *Suicide & Life-Threatening Behavior*, *53*(3), 522–533. https://doi.org/10.1111/sltb.12961

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., & the 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*, *35*(6), 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

Trivedi, R. B., Post, E. P., Sun, H., Pomerantz, A., Saxon, A. J., Piette, J. D., Maynard, C., Arnow, B., Curtis, I., Fihn, S. D., & Nelson, K. (2015). Prevalence, comorbidity, and prognosis of mental health among U.S. veterans. *American Journal of Public Health*, *105*(12), 2564–2569. https://doi.org/10.2105/AJPH.2015.302836

Velikonja, D., Warriner, E., & Brum, C. (2010). Profiles of emotional and behavioral sequelae following acquired brain injury: Cluster analysis of the Personality Assessment Inventory. *Journal of Clinical and Experimental Neuropsychology*, *32*(6), 610–621. https://doi.org/10.1080/13803390903401302

West, A. N., & Weeks, W. B. (2009). Health care expenditures for urban and rural veterans in Veterans Health Administration care. *Health Services Research*, *44*(5), 1718–1734. https://doi.org/10.1111/j.1475-6773.2009.00988.x

Williams, E. C., Fletcher, O. V., Frost, M. C., Harris, A. H., Washington, D. L., & Hoggatt, K. J. (2022). Comparison of substance use disorder diagnosis rates from electronic health record data with substance use disorder prevalence rates reported in surveys across sociodemographic groups in the Veterans Health Administration. *JAMA Network Open*, *5*(6), Article e2219651. https://doi.org/10.1001/jamanetworkopen.2022.19651

Worthen, M. D., & Moering, R. G. (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

Wygant, D. B., Sellbom, M., Gervais, R. O., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. L. (2010). Further validation of the MMPI-2 and MMPI-2-RF Response Bias Scale: Findings from disability and criminal forensic settings. *Psychological Assessment*, *22*(4), 745–756. https://doi.org/10.1037/a0020042

Young, G., Erdodi, L., Giromini, L., & Rogers, R. (2025). Malingering-related assessments in psychological injury: Performance Validity Tests (PVTs), Symptom Validity Tests (SVTs), and invalid response set. *Psychological Injury and Law*, *18*(1), 19–34. https://doi.org/10.1007/s12207-024-09523-6

Received March 12, 2025
Revision received April 28, 2025
Accepted April 30, 2025 ■