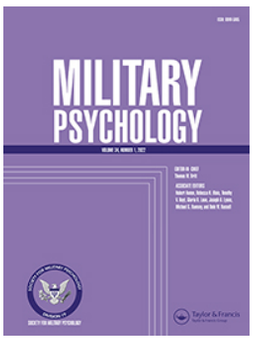

Routledge
Taylor & Francis Group

Military Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hmlp20

Relationship of personality assessment inventory (PAI) over-reporting scales to performance validity testing in a military neuropsychological sample

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

To cite this article: Nicole M. Morris, Paul B. Ingram & Patrick Armistead-Jehle (2022): Relationship of personality assessment inventory (PAI) over-reporting scales to performance validity testing in a military neuropsychological sample, Military Psychology, DOI: 10.1080/08995605.2021.2013059

To link to this article: https://doi.org/10.1080/08995605.2021.2013059

Published online: 29 Mar 2022.

MILITARY PSYCHOLOGY
https://doi.org/10.1080/08995605.2021.2013059

Routledge
Taylor & Francis Group

Relationship of personality assessment inventory (PAI) over-reporting scales to performance validity testing in a military neuropsychological sample

Nicole M. Morris, M.A.[a], Paul B. Ingram, PhD[a,b], and Patrick Armistead-Jehle, PhD, ABPP-CN[c]

[a]Department of Psychological Sciences, Texas Tech University, Lubbock, Texas, USA; [b]Dwight D. Eisenhower Veteran Affairs Medical Center, Eastern Kansas Veteran Healthcare System, Leavenworth, Kansas, USA; [c]Munson Army Health Center, Leavenworth, Kansas, USA

ABSTRACT

This study evaluated the Personality Assessment Inventory's (PAI) symptom validity-based over-reporting scales with concurrently administered performance validity testing in a sample of active-duty military personnel seen within a neuropsychology clinic. We utilize two measures of performance validity to identify problematic performance validity (pass all/fail any) in 468 participants. Scale means, sensitivity, specificity, predictive value, and risk ratios were contrasted across symptom validity-based over-reporting scales. Results indicate that the Negative Impression Management (NIM), Malingering Index (MAL), and Multiscale Feigning Index (MFI) scales are the best at classifying failed performance validity testing with medium to large effects (d = .61–.73). In general, these scales demonstrated high specificity and low sensitivity. Roger's Discriminant Function (RDF) had negligible group differences and poor classification. The Feigned Adult ADHD index (FAA) performed inconsistently. This study provides support for the use of several PAI over-reporting scales at detecting probable patterns of performance-based invalid responses within a military sample. Military clinicians using NIM, MAL, or MFI are confident that those who elevate these scales at recommended cut scores are likely to fail concurrent performance validity testing. Use of the Feigned Adult FAA and RDF scales is discouraged due to their poor or mixed performance.

ARTICLE HISTORY
Received 11 August 2021
Accepted 23 November 2021

KEYWORDS
Personality Assessment Inventory; PAI; Feigning; Military; Over-reporting

What is the public significance of this article?—This study evaluates the classification accuracy of the PAI validity scales in active-duty neuropsychology evaluations and supports the use of some scales.

Detection of content-based invalid responding (i.e., either intentional or unintentional effort by respondents to misrepresent their symptoms, attitudes, or beliefs through inaccurate item responses) is an important process in ensuring accurate psychological assessments (Burchett & Ben-Porath, 2010; Roger & Bender, 2018). Given that test data invalidated by symptom misrepresentation markedly impacts the interpretability of an instrument's clinical scale, it is imperative to ensure only accurate data is incorporated into assessment case formulations (Burchett & Ben-Porath, 2010; Wershba et al., 2015; Wiggins et al., 2012). It is for this reason that assessment inventories typically include validity scales designed specifically to detect response invalidity.

Concerns about adequate detection of invalid responses are pronounced across numerous special populations, including veterans (Ingram et al., 2019; Ray, 2017) whose military records may be utilized during post-discharge evaluations (Worthen & Moering, 2011). Active-duty personnel in the United States (U.S.A.) military can pose a similar challenge, with the detection of invalid responses presenting them as a major challenge in neuropsychological evaluations (Armistead-Jehle & Buican, 2012; Grills & Armistead-Jehle, 2016). Researchers have, for instance, found that performance validity results account for most of the variance in cognitive test scores in military samples, beyond history of concussion, psychological functioning, and demographic variables (Armistead-Jehle et al., 2016). Thus, assessing data validity via performance validity testing is critical when making clinical determinations within a military population. In addition, approximately 20–35% of those with a history of mild traumatic brain injury (mTBI) have evidence of performance validity testing (PVT) failure during their evaluation (Lange et al., 2012; Armistead-Jehle et al., 2018). Such a high rate of PVT failure for those with a history of mTBI is alarming from the perspective of obtaining interpretability psychological testing data, particularly given the increased incidence of mTBI in the military

CONTACT Paul B. Ingram pbingram@gmail.com Texas Tech University, 2810 18th Street, Box 42051, Lubbock, Texas 79410, USA.

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

over the past decade and its general prevalence as an evaluation concern (Defense and Veteran Brain Injury Center [DVBIC], 2019). In short, efforts to better the detection of invalid responses are critical to effective diagnostic practice within U.S.A. military populations and within neuropsychological evaluations conducted on them more specifically.

The Personality Assessment Inventory (PAI; Morey, 1991, 2007) is a widely used measure of personality and psychopathology (Ingram et al., 2020b; Wright et al., 2017), including within neuropsychological settings where 27.5% of the neuropsychologists report utilizing the PAI to make determinations to test validity (Martin et al., 2015). In general, research has demonstrated the PAI's reliability and validity under a variety of diagnostic conditions and evaluative needs common in military population, both active duty and veteran (e.g., Bellet et al., 2018; Ingram et al., 2019; Mozley et al., 2005). The PAI also includes several symptom validity scales designed to detect specific patterns of invalid responses that might influence protocol validity (e.g., over-reporting, under-reporting, and non-content-based responding).

Specific to the current study, within standard PAI scoring, there are several scales aimed to assess over-reporting, including the Negative Impression Management (NIM), Malingering (MAL; L. C. Morey, 1996), and Rogers Discriminant Function (RDF; Rogers et al., 1996) scales. In a meta-analysis of these three, frequently utilized over-reporting scales, Hawes and Boccaccini (2009) found that among criterion classified feigning groups, the NIM and MAL scales were best at discriminating between feigned and honest responding (d = 1.06 and .94, respectively). Conversely, the RDF scale had minimal utility discriminating between criterion-classified groups (d = .31).

In addition to the scales examined by Hawes and Boccaccini (2009), several recently developed supplemental validity scales are also available on the PAI (McCredie & Morey, 2018), including the Multiscale Feigning Index (MFI; Gaines et al., 2013), the Feigned Adult Attention-Deficit Hyperactivity Disorder Scale (FAA; Aita et al., 2018), and the Negative Distortion Scale (NDS; Mogge et al., 2010). Gaines et al. (2013) developed the MFI by summing seven clinical scale scores to detect if the respondents are endorsing higher rates of broadband pathology than is expected in inpatient forensic populations. The NDS was developed using 15 items rarely endorsed items to identify over-reported pathology. Consistent with their initial validations, both NDS and MFI demonstrated utility in detecting invalid responding (Russell & Morey, 2019; Wooley & Rogers, 2015). While FAA showed promise in its initial validation, a subsequent independent study found it had limited evidence for its utility (Harrison et al., 2019). In addition to the mixed results in the detection of feigned ADHD, research on FAA has not examined the utility of the scale to detect invalid cognitive symptoms more broadly (e.g., memory – which is a common area of concern within ADHD; Carlozzi et al., 2015; Theiling & Petermann, 2016). As an alternative to the symptom validity approaches utilized for the other PAI over-reporting scales, the Cognitive Bias Scale (CBS; Gaasedelen et al., 2019) explicitly focuses on cognitive symptom assessment within the PAI. CBS used PVTs to identify scale items, mirroring the methods of highly effective cognitive scales on other broadband measures (e.g., MMPI-2-RF's RBS scale; see Ingram & Ternes, 2016). Subsequently, there have been several studies that have repeatedly supported the CBS's utility against performance-based criterion, including in military populations (Armistead-Jehle et al., 2020; Boress et al., 2021; Tylicki et al., 2021).

Studies validating PAI over-reporting scales have typically focused on comparisons with measures of symptom validity (e.g., Gaines et al., 2012; Mogge et al., 2010). Comparison of PVT and PAI-based Symptom Validity Testing (SVT) outcomes are less well known (see Fokas & Brovko, 2020), despite being related constructs. Indeed, this area of the literature is just starting to grow within research on the PAI. Mooney, Stafford, and Seats (2018) found those with failed PVTs scored higher on NIM, MAL, and RDF, but not enough to aid in determinations of profile invalidity (e.g., they demonstrated statistical, but not clinically meaningful, levels of difference). Conversely, Gaasedelen et al. (2019) found that NIM and MAL could effectively differentiate those passing and failing performance validity testing. Subsequent analyses by Whiteside et al. (2020) supported the work of Gaasedelen and colleagues, indicating those performing at chance levels during performance validity testing are likely to endorse higher rates of psychopathology. Thus, there are mixed results on the effectiveness of some PAI over-reporting SVTs (e.g., NIM and MAL) to detect PVT failure, despite frequent use of these scales in neuropsychological evaluations (Martin et al., 2015).

More broadly than just the PAI, but consistent with the research on PAI SVT functioning (e.g., Mooney et al., 2018; Whiteside et al., 2020), studies comparing SVT and PVT performance have often found similarly mixed results. For instance, Bomyea et al. (2020) investigated the associations between an embedded SVT in the Neurobehavioral Symptoms Inventory and performance validity and concluded that SVT performances were a poor predictor of PVTs. Conversely, Ingram et al. (2020a) evaluated the

relationship between performance validity and symptom validity measures (e.g., over-reporting validity scales on the MMPI-2-RF personality assessment inventory) in an active-duty sample. These authors reported small to medium effect sizes and high specificity for the SVTs performance against PVTs, albeit low sensitivity. Ingram et al. (2020a) concluded that among those who fail broadband personality measure SVTs, there is a good chance that PVTs will also be failed, but many individuals who fail PVTs will not elevate broadband SVTs to a point of invalidity. Taken as a whole, research on the relationship between performance validity testing and SVT measures (such as most of the PAI over-reporting scales) remains limited and an area of needed investigation.

The purpose of this study was to examine the effectiveness of the PAI over-reporting scales designed based on infrequent symptom methodology at detecting failed performance validity testing in an active-duty military population. Given that PVT and SVTs are frequently co-administered within neuropsychological settings, understanding associations between their performance is critical for effective integrative clinical interpretation. Specifically, this study: (i) examines the capacity of the NIM, MAL, RDF, MFI, and FAA scales to differentiate between active-duty military personnel classified based on failed performance validity testing and (ii) calculates sensitivity, specificity, and classification accuracy estimates for each scale. Consistent with prior studies, large effect size estimates were also expected for the NIM and MFI scales whereas medium effects were expected for RDF and MAL. Given mixed results (Aita et al., 2018; Harrison et al., 2019) and a lack of data on FAA's performance in cognitive feigning tasks more broadly, we approached these analyses as exploratory. Consistent with research on previously embedded validity indicators within personality assessment instruments within military populations (e.g., Ingram et al., 2020a; Jones, 2016), it was also hypothesized that at acceptable levels of specificity (>~.90), sensitivity would be limited (< ~.30) for each scale at their recommended cut-scores.

Methods

Participants

Participants were active-duty United States Army service members evaluated in an outpatient neuropsychological clinic. Participants received consecutive patient referrals for neuropsychological evaluation from behavioral health and primary care as the primary sources of referral. Participants were evaluated between 2011 and 2018. The specific referral source of the participants was not coded in the database and is therefore unavailable. Evaluation occurred at a midwestern United States Army Health Center and data was entered into a clinical database at the time of the evaluation, which this study utilized for analysis. However, to be included in the final sample individuals must have completed both PVTs (i.e., Medical Symptom Validity Test [MSVT] and Nonverbal Medical Symptom Validity Test [NV-MSVT]) and were excluded if they were undergoing a medical board (MEB; n = 64) or a Temporary Disability Evaluation (TDE) (n = 4) and/or participants exceeded recommended cut scores of the non-content base invalid responding measures of the PAI (i.e., Inconsistency (INC) and Infrequency (INF); INC > 72 and INF > 75).

Our final sample consisted of 468 (407 men [87.0%], 61 women [13.0%]) active-duty United States Army service members. The predominance of males within the sample is consistent with active-duty composition that falls around 15% in the Army (see United States Government Accountability Office, 2020). Most of the sample had available information on their current rank and the sample included both enlisted (n = 174; 43.6%) and officers (n = 225; 56.4%). The most frequently enlisted ranks were E-4 (n = 29), E-5 (n = 28), E-6 (n = 38), and E-7 (n = 37), while the largest portion of officers held the rank of O-3 (n = 130) or O-4 (n = 43). In general, participants were white (73.5%), 37 years old (SD = 7.7), and had an average education of 15.5 years (SD = 2.4). In terms of diagnosis, 67.3% of the individuals had a history of mTBI and/or concussions and 88.5% were diagnosed with one of the following psychiatric conditions: anxiety disorder (28.3%), any unipolar depressive disorder (e.g., Major Depressive Disorder, Dysthymia/Persistent Depressive Disorder, or Depressive Disorder NOS; 11.0%), posttraumatic stress disorder (10.6%), and both a unipolar depression and PTSD (8.2%). Data was unavailable about the concurrent use of mental health services by participants, or the rate/type of referral made for mental health services for those needing such care who were not previously under such care. See Table 1 for additional demographic information and descriptive cognitive testing of the sample.

Instrumentation

Personality assessment inventory

The PAI (Morey, 1991, p. 1997) is a 344-item self-reported personality assessment instrument that measures protocol validity, clinical presentation, treatment consideration, and interpersonal characteristics. PAI items are answered using a 4-point range (*false, somewhat true, mainly true*, or *very true)*. The validity

Table 1. Demographic and neuropsychological descriptive information.

	Full Sample				0 PVT Failures				Failed Any (1 or 2) PVT(s)			
	(*n* = 468)				(*n* = 326)				(*n* = 142)			
Variable	*n*	*M*	SD	%	*n*	*M*	SD	%	n	M	SD	%
Age	468	37.0	7.7		326	36.8	7.5		142	37.6	8.2	
Years of Education	468	15.5	2.4		326	15.6	2.3		142	15.1	2.4	
Gender (Male)	407			87.0%	283			86.8%	124			87.3%
Ethnicity												
White	344			73.5%	258			79.1%	86			60.6%
African American	80			17.1%	42			12.9%	38			26.8%
Hispanic	30			6.4%	17			5.2%	13			9.2%
Other	14			3.0%	9			2.8%	5			3.5%
History of Concussion/mTBI	315			67.3%	213			65.3%	102			71.8%
Psychiatric Diagnosis												
Depression	51			11.0%	31			960.0%	20			14.2%
PTSD	49			10.6%	24			7.5%	25			17.7%
PTSD & Depression	38			8.2%	18			5.6%	20			14.2%
Anxiety	131			28.3%	93			28.9%	38			27.0%
None	145			30.2%	116			36.0%	24			17.0%
WAIS-IV												
FSIQ	96	106.0	14.0		73	108.2	12.8		23	96.6	13.3	
VCI	96	109.2	13.6		73	111.8	12.9		23	101.0	14.1	
PRI	97	105.1	18.1		74	106.4	19.5		23	99.7	14.5	
WMI	98	102.3	14.2		75	104.0	12.9		23	96.8	15.1	
PSI	99	99.6	13.0		76	102.7	10.8		23	89.5	12.4	
COWAT	445	43.8	10.0		306	45.2	10.2		139	41.0	9.1	
RBANS												
Total	361	95.2	14.4		243	100.0	11.7		118	85.5	14.2	
Immediate Memory	362	96.4	14.8		243	99.4	13.2		119	90.1	16.0	
Visuospatial/Construction	362	100.6	15.6		243	103.0	14.3		119	95.5	16.6	
Language	362	97.2	12.8		243	99.6	12.3		119	92.5	12.8	
Attention	362	94.6	16.7		243	98.9	15.2		119	86.0	16.3	
Delayed Memory	362	92.9	18.0		243	99.5	12.9		119	79.9	19.5	
TMT A	450	43.1	13.3		310	45.6	12.6		140	37.5	13.2	
TMT B	450	45.7	11.2		310	47.8	10.5		140	41.0	11.3	

Note. WAIS-IV = Weschler Adult Intelligence Scale, Fourth Edition. COWAT = Controlled Oral Word Association Test, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status, and TMT A and B = Trail Making Test form A and B. FSIQ = Full Scale IQ, PSI = Processing Speed Index, WMI = Working Memory Index, VCI = Verbal Comprehension Index, and PRI = Perceptual Reasoning Index. COWAT scores are presented as standardized *T*-scores. DD = depressive disorder, PTSD = posttraumatic stress disorder, SUD = substance use disorder, ADHD = attention-deficit hyperactivity disorder.

scales of the PAI cover content related to (i) non-content based invalid responding (i.e., Inconsistency [ICN] and Infrequency [INF]), (ii) content-based under-reporting (i.e., positive distortion; Positive Impression Management [PIM], Defensiveness Index [DEF], and Cashel Discriminant Function [CDF]), and (iii) and content-based over-reporting (i.e., negative distortion; Negative Impression [NIM], Malingering Index [MAL], Rogers Discriminant Functioning [RDF]). For a summary of the over-reporting measures included within the PAI, as well as those developed as supplemental measures, see Table 2. The clinical scales of the PAI utilize *T*-scores ≥70 to indicate significant elevations, and these scales have demonstrated good psychometric properties related to reliability and validity (see L. C. Morey, 1996; Morey, 1991, 2007).

Performance validity testing

To classify individuals into categories based on their performance, two stand-alone performance validity tests [PVT] were administered. For performance validity testing, we utilized the Medical Symptom Validity Test (MSVT; Green, 2004) and Nonverbal Medical Symptom Validity Test (NV-MSVT; Green, 2008). Within our sample, most individuals failed no PVTs (*n* = 349; 74.6%), while approximately one-third participants failed either one (*n* = 66; 14.1%) or two (*n* = 53; 11.3%) PVTs.

Medical symptom validity test

The MSVT (Green, 2004) is a verbal memory-based PVT consisting of 10-word pairs and 5 subtests. Failure of this measure was defined as scoring below the specified cut score on any one of the first three subtests (see Carone, 2009 for review of the MSVT).

Nonverbal medical symptom validity test

The NV-MSVT (Green, 2008) is a non-verbal memory-based PVT consisting of 10 paired images and seven subtests. Failure of this measure was determined per manual instructions. (For a review of the NV-MSVT as a performance validity indicator, see Wagner & Howe, 2010).

Table 2. Overview of PAI validity scales paper.

PAI Validity Scales	Abbreviation	Content Assessed	Citation
Non-Content Based Responding Scales			
Inconsistency scale	INC	Consistency of responding to items with similar content	Morey (1991)
Infrequency scale	INF	Random or atypical responding	Morey (1991)
Negative Distortion Scales			
Negative Impression scale	NIM	Exaggerated negative responding	Morey (1991)
Malingering Index	MAL	Simulation of severe psychopathology	L. C. Morey (1996)
Rogers Discriminent Function Index	RDF	Intentional simulation of psychopathology	Roger et al. (1996)
Multiscale Feigning Index	MFI	Extreme representation of pyschopathology	Gaines et al. (2013)
Feigned Adult ADHD Index	FAA	Feigned ADHD sympomology	Aita et al. (2018)
Positive Distortion Scales			
Positive Impression Scale	PIM	Exaggerated positive responding or denial of minor faults	Morey (1991)
Defensiveness Index	DEF	Positive impression or defensive respdoning	L. C. Morey (1996)
Cashel Discriminant Function	CDF	Positive defensive responding	Cashel et al. (1995)

Procedures and planned analysis

This investigation received IRB approval from Madigan Army Medical Center. Prior to analysis, which occurred in 2021, participants were planned for exclusion if their scores on INF or ICN exceeded the recommended values in the PAI's technical manual (Morey, 1991, 2007). The remaining participants were grouped into two categories (pass all PVTs without failure or failed one or more PVTs) based on PVT performance. Independent *t*-tests were used to examine between-group differences for each of the following PAI validity scales: NIM, RDF, MAL, FAA, and the MFI. NDS was not included because item-level scores required for calculating this scale were not available within this retrospective database. CBS was not calculated as it also required item-level scores. Hedge's *g* was used to estimate the magnitude of the effect for differences between groups to account for differences in sample size. We classified between-group differences utilizing Cohen's (1988) recommendations of small (.5 > *g* > .2), medium (.8 > *g* ≥ .5), and large (.8 ≥ *g*) effects and 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.). Receiver operator curve (ROC) characteristics were calculated for each over-reporting scale, along with specificity, sensitivity, hit rate, positive predictive value (PPV), negative predictive value (NPV) and relative risk ratios (RRR). A board-certified neuropsychologist (the third author of this paper, Dr. Armistead-Jehle) provided clinical interpretation and diagnostic formulations based on testing data, including classification of head injury severity and diagnosis as they are reported within the participant demographics.

Results

Descriptives for each scale, along with results of independent t-tests are provided in Table 3. As it typical in feigning research (feigning groups typically demonstrate larger standard deviations; see Hawes & Boccaccini, 2009; Ingram & Ternes, 2016), there were variations in homogeneity across groups on several of the validity scales of the PAI within this study (i.e., NIM, MAL, and FAA). T-test results presented for the NIM, MAL, and FAA scales utilize an assumption of non-equality of variances. Results suggest that the NIM, MAL, FAA, and MFI over-reporting scales differ significantly between individuals across PVT failure groups. No such differences were observed on RDF. Individuals failing PVTs scored significantly higher than those who did not fail any PVTs on the NIM, MAL, MFI and FAA scales. The magnitude of this

Table 3. Differences in the PAI scales according to extra-test grouping criteria.

			Number of PVT Failures								
	Full Sample (*n* = 468)		Failed None (*n* = 326)			Failed Any (*n* = 142)					
Scale	*M*	*SD*	*M*	*SD*	% ≥ *RCS*	*M*	*SD*	% ≥ *RCS*	*Levine's Test F-value (df=)*	*t*	*g*
NIM	53.56	11.36	51.40	9.07	< . 01%	59.76	14.65	2.8%	34.12*	−5.85*	**.73**
MAL	52.68	11.28	50.93	9.39	< . 01%	57.88	14.40	8.5%	20.82*	−4.92*	**.61**
RDF	47.22	9.89	47.07	9.52	< . 01%	47.56	10.93	5.6%	3.40	−.47	.06
FAA	1.99	4.32	1.62	2.31	52.5%	3.64	5.73	73.9%	39.48*	−3.75*	**.55**
MFI	49.24	10.56	47.03	8.60	1.5%	54.89	9.73	2.8%	2.07	−8.31*	**.73**

Note. PVT = performance validity test; NIM = Negative Impression scale; RDF = Rogers Discriminant Function scale; MAL = Malingering Index scale; FAA = Feigned Adult ADHD index; MFI = Multiscale Feigning Index; *M* = mean; SD = standard deviation; % ≥ RCS = percentage of sample at or above recommended cut score. *F* = Levene's test *F*-value; * = $p < .01$; *t* = *t* statistic; *g* = Hedge's g effect size.

effect was consistent across the NIM, MAL, MFI, and FFA scales and fell within the medium effect range (g = .55–.73). RDF had negligible effect sizes (g = .06). Broadly, patterns of significance and the observed magnitudes of effect suggested that NIM, MAL, MFI, and FAA were best at discerning those who were likely to fail PVTs.

Next, ROC analyses estimated Area Under Curve [AUC][1] along with classification statistics (i.e., sensitivity, specificity, hit rate, positive predictive power [PPP], and negative predictive power [NPP]) for each PAI scale. Results of AUC analyses were medium (NIM and MAL) to large (MFI and FAA) in effect for the scales, which demonstrated significant between-group differences: NIM = .679 (medium effect, standard error [SE] = .028, 95% confidence interval [CI] = .625 to .734), MAL = .631 (small effect, SE = .029; 95% CI = .574 to .688), MFI = .724 (large effect, SE = .025, 95% CI = .675 to .773), and FAA = .653 (large effect, SE = .027, 95% CI = .600 to .707). In general, results indicated greater than chance, but less than ideal, classification for these three scales. For RDF (AUC = .514 [SE = .030, 95% CI = .456 to .572]), results fell in the negligible range.

More specific classification information is provided through the sensitivity, specificity, PPP, and NPP estimates for each of the five scales (see Meehl & Rosen, 1995), located in Table 4. For NIM, MAL, and MFI, we generally observed strong specificity (.96–.99) and PPP (.68–.96) with rather weak sensitivity (.03–.08) and acceptable NPP (.61 to .62) at recommended cut scores. For FAA at its recommended cut score, we observed moderate specificity (.50), weak PPP (.10) and sensitivity (.22), and acceptable NPP (.72). Recommended cut scores are as follows: 92 T for NIM (Morey, 1991, 2007), 72 T (Raw score of 3 in clinical normative samples) for MAL (L. C. Morey, 1996, p. 126), 76 T for MFI (Gaines et al., 2013), and .60 for FAA (Aita et al.). At recommended values, individuals who are classified as having significantly elevated most of these scales (i.e., NIM, MAL, and MFI) are also more likely to have failed at least some PVT testing (see Table 3). For FAA, the high rate of individuals exceeding the recommended cut value even in those without PVT failures (52.5%) makes the increased probability of PVT failure less clinically meaningful.

Table 4. Classification estimates for the PAI over-reporting validity scales.

Scale	Sensitivity	Specificity	OCC	PPP	NPP
NIM					
≥ 99	.03	1.00	.61	.86	.61
≥ 92	**.06**	**.99**	**.62**	**.80**	**.61**
≥ 88	.06	.99	.62	.80	.61
≥ 84	.07	.98	.62	.75	.61
≥ 80	.13	.98	.64	.81	.63
≥ 76	.18	.97	.65	.78	.64
≥ 72	.25	.96	.67	.80	.66
≥ 68	.30	.91	.67	.70	.66
≥ 64	.33	.86	.65	.62	.66
RDF					
≥ 82	0.00	1.00	0.60	0.00	0.60
≥ 72	.03	.99	.61	.74	.61
≥ 70	**.03**	**.96**	**.59**	**.37**	**.60**
≥ 68	.03	.96	.59	.37	.60
≥ 66	.05	.96	.59	.42	.60
≥ 64	.08	.95	.60	.51	.61
≥ 62	.12	.93	.60	.52	.61
≥ 60	.21	.91	.63	.60	.63
≥ 58)	.26	.84	.61	.53	.63
≥ 56	.29	.77	.58	.46	.62
MAL (raw score)					
≥ 4	.04	1.00	.61	1.00	.61
≥ 3	**.08**	**.99**	**.63**	**.96**	**.62**
≥ 2	.27	.91	.66	.67	.65
≥ 1	.61	.59	.60	.50	.69
≥ 0	1.00	.00	.40	.40	.69
FAA					
≥ 6	.04	.95	.76	.15	.80
≥ 5	.05	.91	.74	.13	.79
≥ 4	.07	.90	.73	.15	.79
≥ 3	.11	.84	.69	.14	.79
≥ 2	.12	.79	.66	.13	.78
≥ 1	.17	.64	.55	.11	.76
≥ .80	.19	.58	.50	.10	.74
≥ .60	**.22**	**.50**	**.44**	**.10**	**.72**
≥ .40	.24	.35	.33	.08	.65
MFI					
≥ 76	**.03**	**.99**	**.61**	**.68**	**.61**
≥ 74	.04	.99	.61	.73	.61
≥ 72	.05	.99	.61	.76	.61
≥ 70	.06	.98	.61	.68	.61
≥ 68	.11	.97	.63	.75	.62
≥ 66	.16	.95	.64	.70	.63
≥ 64	.21	.93	.64	.66	.64
≥ 62	.24	.89	.63	.59	.64
≥ 60	.28	.86	.63	.57	.64

Notes. NIM = Negative Impression scale; RDF = Rogers Discriminant Function scale; MAL = Malingering Index scale; FAA = Feigned Adult ADHD index; MFI = Multiscale Feigning Index; OCC = overall correct classification; PPP = positive predictive power; NPP = negative predictive power.

Discussion

This study investigated the utility of several PAI over-reporting indicators in detecting PVT failure in a sample of active-duty U.S. Army personnel, addressing a needed area of study for the PAI (Fokas & Brovko, 2020; Martin et al., 2015). Specifically, we evaluated how well scores on NIM, MAL, RDF, FAA, and MFI were able to differentiate between those who passed and failed performance validity testing. Results from this study broadly suggest three distinct findings. First, among the PAI over-reporting scales, the NIM, MAL, and MFI scales appear to function most effectively at differentiating failure using PVT testing as the criterion while FAA offers mixed evidence of effectiveness. Second, the magnitude of effects observed across PVT failure groups ranged from medium to large, suggesting discernable score differences between the groups on these scales. Third, using

the recommended cut-scores for NIM, MAL, and MFI leads to high specificity and low sensitivity, consistent with patterns observed in other self-reported personality assessments. FAA produced moderate specificity and low sensitivity. Several aspects of these results warrant additional discussion.

Scale effectiveness observed on over-reporting indicators (i.e., NIM, MFI, MAL) appears effective at discriminating PVT failure within an active-duty military sample seen within a neuropsychological clinic, when utilizing mean difference approaches. Given the observed rates of low sensitivity and high specificity at recommended cut scores (and at scores substantially lower than recommend cut values), clinicians should have confidence in utilizing these scales to make determinations about profile validity, particularly in concluding an individual is likely to fail performance validity testing when they exceed recommended scale cut-scores. While sensitivity rates were less than ideal, the high specificity remained evident in those who failed a single PVT, as well as those who failed two or more which support the rule-in approach to assessing effort (e.g., failure on this scale is likely to indicate suboptimal performance on PVTs). These several over-reporting indicators of the PAI offer a level of specificity that is generally consistent with the performance of embedded indicators of over-reporting on other popular broadband personality instruments generally (see Sharf et al., 2017), as well as within comparable military neuropsychology clinics (Ingram et al., 2020a). Still, the low sensitivities of the PAI SVTs indicate that these scales cannot replace PVTs in clinical evaluation. We also note that failure on a PVT or SVT does not, however, infer feigning or malingering alone, as such a determination requires an additional burden of evidence (see Sweet et al., 2021).

The MFI scale had low sensitivity and high specificity; however, a reduced cut-score may be necessary in military neuropsychological evaluations to perform similarly to its initial validation ($T = 76$; Gaines et al., 2013). Given that the MFI was developed for use with forensic populations, a different cut score within a neuropsychological evaluation sample is not unexpected. Based on performance observed within our study, a cut-score of $T = 64$ (rather than $T = 76$) most closely approximates the balance of modest sensitivity (~.3) and high specificity (~.9) common in over-reporting measures within neuropsychological settings (Armistead-Jehle et al., 2020; Gaasedelen et al., 2019; Gervais et al., 2007). In summation, the MFI provides comparable utility to the scales discussed above and, perhaps, even better utility given its higher sensitivity when utilizing a lower cut score.

Accordingly, the strengths and concerns about the use of most PAI indicators (i.e., NIM, MAL, and MFI) mirror those of other embedded indicators on personality measures. While we can be confident about our ability to classify individual scoring above the respective cut-offs as likely to fail external PVTs, we have limited utility in detecting all of those that will fail PVTs when using recommended cut scores. Many of the scales demonstrated sensitivities that were extremely low. An interesting caveat to these findings is the observation of minimal improvement or change in classification accuracy if lower cut-scores were to be implemented across each of the scales that demonstrated between-group effectiveness. This pattern in response classification effectiveness occurred during both ROC analyses, suggesting that use of the recommended cut scores is appropriate when concerned with specificity. Individuals failing PVTs are likely to be classified correctly when using recommended PAI over-reporting scale cut scores and these same individuals are also likely to have consistent and moderate differences across cognitive and neurocognitive assessment performances compared to those without any PVT failures, providing collateral support for consistent differences identified on the PAI over-reporting measures.

Although designed to assess ADHD specifically, the FAA's focus on memory impairment (e.g., working and short term) reflects the broad deficits, which occur both in ADHD and other neuropsychological conditions (Carlozzi et al., 2015; Theiling & Petermann, 2016). FAA demonstrated the ability to differentiated between groups of individuals passing and failing performance-based memory tasks, suggesting FAA assesses attention and concentration deficits within the scope of neuropsychological practice. These symptoms also align with frequent referral questions among active-duty United States Army service members evaluated in an outpatient neuropsychological clinic. While FAA demonstrated mean group differences and large AUC value effects, FAA scores had poor specificity (.50) and sensitivity (.22) at recommended cut values. Further, score ranges varied so widely from its validation that alternative cut score are unlikely to generalize. Thus, FAA does not provide consistent evidence of its utility in this neuropsychological clinic based on failed performance testing (much of which focuses on memory function) within this study. In conjunction with research by Harrison et al. (2019), which note that the FAA's initial utility (Aita et al., 2018) were not replicated in other ADHD assessment research, caution surrounding FAA's effectiveness is warranted at this time.

Taken together, there are several strengths to the current study. Clinicians can have improved confidence in how they may utilize various PAI over-reporting indicators as collateral measures of response validity within military neuropsychological evaluations. Further, base rate of PVT failure within this study were also consistent with what is expected in neuropsychological evaluations (Larrabee et al., 2009). However, the current study is also not without limitations. Participants administered different sets of PVTs may produce different patterns of performance on these same scales. Said another way, the underlying psychometric function of the administered PVTs may impact group classifications within this study (e.g., the N-MSVT and MSVT's ability to correctly assessed failed cognitive performance may differ from other PVTs). However, the widely validated nature of the administered PVTs within this study help to assure that misclassification of likely PVT failure on other measures, while possible, is not the probable outcome for any participant. Moreover, clinicians may have increased confidence in our findings when viewing groups with greater numbers of failed concurrent PVTs as this suggests a general convergence of evidence about test performance. Additionally, the demographics of our study are composed primarily of White males and, as such, diversity factors such as sex, gender, ethnicity, and race may also be important to consider when applying results within this study to dissimilar individuals. We did not have access to data on if participants were currently receiving mental healthcare, and it will be fruitful for future studies to examine if differences in PVT/SVT failure differ as a function of treatment engagement. This study also utilized PAI scale total scores (and scales calculated using total scores) to conduct analyses. Item-level differences were not evaluated. Future studies may wish to use item response theory in their evaluation of validity-scale utility. Differences in sample sizes may have influenced statistical significance in independent *t*-tests; however, group sizes reflect naturalistic base rate of PVT failure within this sample. Thus, findings are likely generalizable to similar populations/settings, but clinicians should consider base rates (as well as demographics) within their setting while contextualizing findings of this study. Lastly, while classification estimates within this study are consistent with what is typically observed in over-reporting studies on personality inventories (e.g., Tylicki et al., 2021), continuous refinement of scales is needed to improve positive and negative predictive power for best proxy clinical determination accuracy.

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).

Disclosure statement

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

ORCID

Nicole M. Morris, M.A. http://orcid.org/0000-0001-6565-9816
Paul B. Ingram, PhD http://orcid.org/0000-0002-5409-4896
Patrick Armistead-Jehle, PhD, ABPP-CN http://orcid.org/0000-0002-6784-1432

Data availability statement

Due to the nature of this research and IRB stipulation, participants of this study did not agree for their data to be shared publicly, so supporting data is not publicly available.

References

Aita, S. L., Sofko, C. A., Hill, B. D., Musso, M. W., & Boettcher, A. C. (2018). Utility of the personality assessment inventory in detecting feigned Attention-Deficit/Hyperactivity Disorder (ADHD): The feigned adult ADHD index. *Archives of Clinical Neuropsychology*, *33*(7), 832–844.

Armistead-Jehle, P., & Buican, B. (2012). Evaluation context and Symptom Validity Test performances in a US Military sample. *Archives of Clinical Neuropsychology*, *27*(8), 828–839.

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

Armistead-Jehle, P., Cooper, D. B., & Vanderploeg, R. D. (2016). The role of performance validity tests in the assessment of cognitive functioning after military concussion: A replication and extension. *Applied Neuropsychology: Adult*, *23*(4), 264-273.

Armistead-Jehle, P., Ingram, P. B., & Morris, N. M. (2020). Personality assessment inventory cognitive bias scale: Replication in an active duty 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.

Bomyea, J., Jurick, S. M., Keller, A. V., Hays, C. C., Twamley, E. W., & Jak, A. J. (2020). Neurobehavioral symptom validity and performance validity in veterans: Evidence for distinct outcomes across data types. *Applied Neuropsychology: Adult*, *27*(1), 62–72. https://doi.org/10.1080/23279095.2018.1480484

Boress, K., Gaasedelen, O. J., Croghan, A., Johnson, M. K., Caraher, K., Basso, M. R., & Whiteside, D. M. (2021). Replication and cross-validation of the personality assessment inventory (PAI) cognitive bias scale (CBS) in a mixed clinical sample. *The Clinical Neuropsychologist* Advanced Onine Publication. https://doi.org/10.1080/13854046.2021.1889681

Burchett, D. L., & Ben-Porath, Y. S. (2010). The impact of over-reporting on the MMPI-2-RF substantive scale score validity. *Assessment*, *17*(4), 497–516. https://doi.org/10.1177/1073191110378972

Carlozzi, N. E., Kirsch, N. L., Kisala, P. A., & Tulsky, D. S. (2015). An examination of the Wechsler adult intelligence scales, fourth edition (WAIS-IV) in individuals with complicated mild, moderate and severe traumatic brain injury (TBI). *The Clinical Neuropsychologist*, *29*(1), 21–37. https://doi.org/10.1080/13854046.2015.1005677

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

Cashel, M. L., Rogers, R., Sewell, K., & Martin-Cannici, C. (1995). The Personality Assessment Inventory (PAI) and the detection of defensiveness. *Assessment*, *2*, 333–342.

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

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.

Defense and Veteran Brain Injury Center Annual Report. (2019). *Annual Report 2019*. Defense and Veterans Brain Injury Center.

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 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. (2008). Manual for the Nonverbal Medical Symptom Validity Test. Edmonton, Canada: Greens Publishing.

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

Grills, C. E., & Armistead-Jehle, P. (2016). Performance validity test and neuropsychological assessment battery screening module performances in an active duty sample with a history of concussion. *Applied Neuropsychology: Adult*, *23*(4), 295–301. https://doi.org/10.1080/23279095.2015.1079713

Harrison, A. G., Harrison, K. A., & Armstrong, I. T. (2019). Discriminating malingered attention deficit hyperactivity disorder from genuine symptom reporting using novel personality assessment inventory validity measures. *Applied Neuropsychology: Adult*, 1–13.

Harrison, A. G., Lee, G. J., & Suhr, J. A. (2021). Use of performance validity tests and symptom validity tests in assessment of specific learning disorders and attention-deficit/hyperactivity disorder. In K. B. Boone (Ed.), *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective*, 525–558. The Guilford Press.

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

Ingram, P. B., Golden, B. L., & Armistead-Jehle, P. (2020a). Evaluating the minnesota multiphasic personality inventory-2-Restructured Form (MMPI-2-RF) over-reporting scales in a military neuropsychology clinic. *Journal of Experimental and Clinical Neuropsychology*, *42*(3), 263–273. https://doi.org/10.1080/13803395.2019.1708271

Ingram, P. B., Schmidt, A. T., Bergquist, B., & Currin, J. M. (2020b). 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*. Advance online publication. https://doi.org/10.1037/tep0000348

Ingram, P. B., Tarescavage, A. S., Ben-Porath, Y. S., & Oehlert, M. E. (2019). Patterns of MMPI-2-restructured form (MMPI-2-RF) validity scale scores observed across veteran affairs settings [measurement-based care and psychological assessment in mental health services]. *Psychological Services*. Advanced Onine Publication. https://doi.org/10.1037/ser0000339

Ingram, P. B., & Ternes, M. S. (2016). The detection of content-based invalid responding: A meta-analysis of the MMPI-2-Restructured Form's (MMPI-2-RF) over-reporting validity scales. *The Clinical Neuropsychologist*, *30*(4), 473–496. https://doi.org/10.1080/13854046.2016.1187769

Jones, A. (2016). Cutoff scores for MMPI-2 and MMPI-2-RF cognitive-Somatic validity scales for psychometrically defined malingering groups in a military sample. *Archives of Clinical Neuropsychology*, *31*(7), 786–801. https://doi.org/10.1093/arclin/acw035

Lange, R. T., Pancholi, S., Brickell, T. A., Sakura, S., Bhagwat, A., Merritt, V., & French, L. M. (2012). Neuropsychological outcome from blast versus non-blast: mild traumatic brain injury in US military service members. *Journal of the International Neuropsychological Society*, *18*(3), 595–605.

Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *29*(4), 364–373. https://doi.org/10.1093/arclin/acu019

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

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, *52*(3), 194.

Mogge, N. L., Lepage, J. S., Bell, T., & Ragatz, L. (2010). The negative distortion scale: A new PAI validity scale. *Journal of Forensic Psychiatry and Psychology*, *21*(10), 77–90. https://doi.org/10.1080/14789940903174253

Mooney, S. R., Stafford, J., & Seats, E. (2018). Medical evaluation board involvement, non-credible cognitive testing, and emotional response bias in concussed service members. *Military Medicine*, *183*(11–12), e546–e554.

Morey, L. C. (1991). *The personality assessment inventory professional manual*. Psychological Assessment Resources.

Morey, L. C. (1996). *An interpretive guide to the personality assessment inventory*. Psychological Assessment Resources.

Morey, L. C. (2007). *The personality assessment inventory professional manual* (2nd ed.). Psychological Assessment Resources.

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.

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

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

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

Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detectionof feigned mental disorders on the Personality Assessment Inventory: Adiscriminant analysis. Journal of Personality Assessment, *67*, 629–640

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

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

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. https://doi.org/10.1007/s10862-017-9590-1

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.

Theiling, J., & Petermann, F. (2016). Neuropsychological profiles on the WAIS-IV of adults with ADHD. *Journal of Attention Disorders*, *20*(11), 913–924. https://doi.org/10.1177/1087054713518241

Tylicki, J. L., Rai, J. K., Arends, P., Gervais, R. O., & Ben-Porath, Y. S. (2021). A comparison of the MMPI-2-RF and PAI overreporting indicators in a civil forensic sample with emphasis on the response bias scale (RBS) and the cognitive bias scale (CBS). *Psychological Assessment*, *33*(1), 71–83. https://doi.org/10.1037/pas0000968

United States Government Accountability Office (2020, May). *Female active-duty personnel: Guidance and plans needed for recruitment and retention efforts*. Report to congressional Committees. https://www.gao.gov/assets/gao-20-61.pdf

Wager, J.G., & Howe, L.L.S. (2010). Nonverbal medical symptom validity test: Try faking now! *Applied Neuropsychology, 17*, 305-309.

Wershba, R. E., Locke, D. E., Lanyon, R. I., & Lanyon, R. I. (2015). Analysis of minnesota multiphasic personality inventory-2 restructured form response bias indicators as suppressors or moderators in a medical setting. *Psychological Assessment*, *27*(11), 733–737. https://doi.org/10.1037/a0038802

Whiteside, D. M., Hunt, I., Choate, A., Caraher, K., & Basso, M. R. (2020). Stratified performance on the test of memory malingering (TOMM) is associated with differential responding on the personality assessment inventory (PAI). *Journal of Clinical and Experimental Neuropsychology*, *42*(2), 131–141.

Wiggins, C. W., Wygant, D. B., Hoelzle, J. B., & Gervais, R. O. (2012). The more you say the less it means: Overreporting and attenuated criterion validity in a forensic disability sample. *Psychological Injury and Law*, *5*(3), 162–173.

Wooley, C. N., & Rogers, R. (2015). The effectiveness of the personality assessment inventory with feigned PTSD: An initial investigation of resnick's model of malingering. *Assessment*, *22*(4), 449–458. https://doi.org/10.1177/1073191114552076

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), 187–216. https://doi.org/10.1007/s12207-011-9115-2

Wright, C. V., Beattie, S. G., Galper, D. I., Church, A. S., Bufka, L. F., Brabender, V. M., & Smith, B. L. (2017). Assessment practices of professional psychologists: Results of a national survey. Professional Psychology: Research and Practice, *48*(2), 73.